GILLIS

AM 1946 gi



BOSTON UNIVERSITY GRADUATE SCHOOL

Thesis

THE THEORY OF RANGE

AND ITS APPLICATION TO QUALITY CONTROL

by

Sister Catherine Josephine Gillis, S.N.D.

(A. B., Emmanuel College, 1942)

submitted in partial fulfilment of the requirements for the degree of

Master of Arts.

(A. J., control tensor, 1062)

OUTLINE

Approved

First Reader. Elwess. Mode.

Professor of Mathematics

Second Reader. Ralph Professor of Mathematics

Distribution of the largest and smallest values i

PART II. APPLICATION OF RUSSE TO QUALITY COMMON . . . 32

Approved

OUTLINE

	PAGE
INTRODUCTION. NATURE OF RANGE AND SCOPE OF THE THESIS	1
Nature of Range	. 1
Definition	. 1
Advantages and disadvantages	. 2
Relation to σ	. 2
Scope of the thesis	. 3
Purpose of the thesis	. 3
Division of the material	. 3
PART I. THEORY OF RANGE	. 5
Some special distributions	. 5
Distribution of any k-order statistics	. 5
Distribution of the largest and smallest values in	
a sample	. 8
Distribution of range	. 9
Moments of range	. 9
Deriving the equations	. 10
Constructing the tables	. 18
Percentage limits for distribution of range	. 26
Method of computation	. 26
Conclusion	. 31
PART II. APPLICATION OF RANGE TO QUALITY CONTROL	. 32
Quality control technique	. 32
Nature of quality control	. 32

OUTLINE

							510			. 0											
	á			*				*		4	×			4							
	4	+																			
	4												+		. 7			ij	eIs		
		6			0	,				4.											
													,								
					,				9.		٠										
	ø		,														ilo				
	a	٠		q		4															
								an			d i										
		,												,			a.p.				
												¥				ld		it		a.	
	a																				
																				eu(

•		iv
		PAGE
The control chart	•	. 33
Methods of determining σ		. 35
Application of the technique		. 38
The sample		. 38
Statistics of the sample		
Computation of σ		
Control limits		. 39
Conclusion		. 45
BIBLIOGRAPHY		. 48

		*	*			*			a		Janes loutnes adT	
	a					-				-	Methods of determinings	
,					v						Application of the technique.	
			.4.		2						The sample	
					,			*			eigmes end to soldaidade	
											Computation of T	
				ě								
							,					

LIST OF TABLES

TABLE	P.	AGE
I.	Summary of Constants of Distribution of Range	28
II.	Ratio of Mean Range to Standard Deviation	37
III.	Observations of 204 Measurements of Insulation	
	Resistance	40
IV.	Statistical Measures for Distribution of Insula-	
	tion Resistance Measurements Divided into	
	Subgroups of 4 Units Each	41
٧.	Probability Limits for Distribution of Range	46

LIST OF TABLES

Summary of Constants of Distribution of Hange	.I
not bely and and to Standard Deviation	II.
Observations of 204 Measurements of Insulation	.III
Resistance	
-sluent to not such the for Dietribution of Insula-	.VI
ojni čebiviú sjamerusseM eonsjaiseR nolj	
Subgroups of 4 Units Sach	
Probability Limits for Distribution of Range	. 7

LIST OF FIGURES

FIGUE	RE				PAGE
1.	Curve of Distribution, $y = \varphi(x) \dots$.			•	14
2.	Mean of Range for n = 2 to 1000			0	20
3.	Standard Deviation of Range for n = 2 to 1000				21
4.	Standard Deviation of Range for n = 2 to 20 .				25
5.	β_1 of Range for n = 2 to 20		•	•	27
6.	β_2 of Range for n = 2 to 20	•		•	27a
7.	Control Chart for Standard Deviations	•	•		44
8.	Control Chart for Ranges		•		44

a anaple"; 2 ". . . the difference between the highest record

the singlest massible because of a group of measures. It

JL. B. C. Pippett, The Mathods of Statistics (nescend edition, revised; London: Williams and Margate, Ltd., 1937)

Analysia enlarged edition; New York: Harcourt, Brane and

Mr. S. Imarson, The Arnication of Statistical Methods to Industrial Standard matth and Quality Control

SERVICE TO TRAIL

			2	0 0		Ourve of Distribution, $y = \varphi(x)$	1.
		•				Meen of Renge for n = 2 to 1000	S.
			٠		ot s = n	ToT sandard Deviation of Hange for	3.
	*		*	. os	ot S = n	Standard Deviction of Hange for	4
27	*					Of of Range for n = 2 to 20 .	5.
27a						(3 of Range for n = 2 to 20 .	0
					. anoit	. Control Chart for Standard Devis	7.
						Control Chart for Ranges	.8

INTRODUCTION

NATURE OF RANGE AND SCOPE OF THE THESIS

Statistics, though not a new science, is a developing one. Quite naturally, then, many of its measures are still under investigation. One such measure, the theory of which has been recently developed and is, in fact, still in process of development, is the statistical measure of dispersion, the range.

I. NATURE OF RANGE

<u>Definition</u>. The range has been defined as "...
the difference between the greatest and smallest members of
a sample"; 1 "... the difference between the highest recorded score and the lowest recorded score"; 2 "... the difference between the maximum and minimum observations". 3 Representing the distance between extreme observations, it is
the simplest possible measure of a group of measures. It

L. H. C. Tippett, The Methods of Statistics (second edition, revised; London: Williams and Norgate, Ltd., 1937), p.31.

Analysis (enlarged edition; New York: Harcourt, Brace and Company, 1941), p. 80.

³E. S. Pearson, The Application of Statistical Methods to Industrial Standardisation and Quality Control (London: British Standards Institution, 1935), p. 109.

INTRODUCTION

NATURE OF TANGE AND SCORE OF THE THESIS

Statistics, though not a new acience, is a developing one. Quite naturally, then, many of its measures are still under investigation. One such measure, the theory of which has been recently developed and is, in fact, still in process of development, is the statistical measure of dispersion, the range.

I. NATURE OF BANCE

Definition. The range has been defined as "...

the difference between the greatest and smallest members of
a sample"; "... the difference between the highest recorded score and the lowest recorded score"; "... the difference between the maximum and minimum observations". Representing the distance between extreme observations, it is
the simplest possible measure of a group of measures. It

L. H. C. Tippett, The Methods of Statistics (second edition, revised; London: Williams and Morgate, Ltd., 1937), p. 31.

C. H. Richardson, An introduction to Statistical Analysis (enlarged edition; New York: Harcourt, Erace and Company, 1941), p. 80.

Asthods to Industrial Standardisation and Quality Control (London: British Standards Institution, 1935), p. 109.

gives a comprehensive value of the data in that it includes the limits within which all of the items occur. 4 Hence it appears to be the most natural index of dispersion.

Advantages and disadvantages. Nevertheless, it has been little used for purposes of comparison. The extreme ease with which it may be calculated and its very obvious interpretation which have led to its use in many industrial problems, are frequently more than offset by certain serious objections. Determined by only the two extreme measures, it tells nothing of the form of the distribution within the range. A symmetrical and a J-type frequency curve might have the same value for the range. It tells nothing about the concentration of the measures about the center. If either one (or both) of the extremes is an unusual occurrence it may have quite a disproportionate effect on the range.

Relation to 6. Moreover, the range varies for samples of different sizes taken from the same population, being smaller for smaller samples. For a normal population,

Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc., 1941), p. 236.

⁵G. Undy Yule and M. G. Kendall, <u>An Introduction to the Theory of Statistics</u> (eleventh edition, revised; London: Charles Griffin and Company, Ltd., 1937). p. 134.

⁶Richardson, loc. cit.

gives a comprehensive value of the data in that it includes the limits within which all of the items occur. Hence it appears to be the most natural index of dispersion.

Advantages and disadvantages. Nevertheless, it has been little used for jurposes of comparison. The extreme sace with which it may be calculated and its very obvious interpretation which have led to its use in many industrial problems, are frequently more than offset by certain serious objections. Determined by only the two extreme measures, it tells nothing of the form of the distribution within the renge. A symmetrical and a J-type frequency curve might have the same value for the range. It tells nothing about the concentration of the measures about the denter. It is nothing about the concentration of the measures about the denter. It is an unusual occurrence either one (or both) of the extremes is an unusual occurrence it may have quite a disproportionate effect on the range.

Relation to or Moreover, the range varies for samples of different sizes taxen from the same population, being smaller for sma

General Statistics (New York: Prentice-Hall, LEC., 1941), p. 236.

⁵g. Undy Yule and M. G. Kendall, an introduction to the Theory of Statistics (eleventh edition, revised; London: Coarles Griffin and Company, Ltd., 1937). p. 134.

Oklohardson, loc. olt.

however, the mean range over k samples (k large) bears a fixed relationship to the more common measure of variation, the standard deviation of the population. Nor does this ratio seem to be very sensitive to moderate changes in the form of distribution, so that it can be said that, in general, when the population form and sample size are constant, the mean range is proportional to the standard deviation. It is from this property of the range, as well as from the fact that it is easily computed and easily understood, that its utility arises, its chief use up to the present being in its application to quality control.

II. SCOPE OF THE THESIS

Purpose of the thesis. Because of the growing importance of this phase of statistical work, a discussion of range theory seemed justifiable. How this theory may be utilized to simplify the numerical computations in control chart analysis has also been demonstrated.

<u>Division of the material</u>. This study, then, has been divided into two parts. Part I treats of the theory of range that has been developed up to the present. An attempt

⁷H. A. Freeman, <u>Industrial Statistics</u> (New York: John Wiley & Sons, Inc., 1942), p. 131.

⁸Tippett, loc. cit.

however, the mean range over k samples (k large) bests a fixed relationship to the more couron measure of variation, the standard deviation of the population. Nor does this ratios seem to be very sensitive to moderate changes in the form of distribution, so that it can be said that, in general, when the population form and sample size are constant, the mean range is proportional to the standard deviation. It is from this property of the range, as well as from the fact that it is easily confuted and easily understood, that its utility arises, its order use up to the present being in its application to quality control.

II. SUOPE OF THE THESE

Purpose of the thesis. Because of the growing importance of this phase of statistical work, a discussion of range theory seemed justifiable. How this theory may be utilized to simplify the numerical computations in control chart analysis has also been demonstrated.

Division of the material. This study, then, has been divided into two parts. Part I treats of the theory of range that has been developed up to the present. An attempt

⁷H. A. Freeman, Inquatrial Statistics (New York: John Wiley & Sons, Inc., 1942), p. 131.

Stippett, loc. oit.

has been made to collect all available material dealing with the distribution of range in general, and in particular with the distribution of range when the samples are drawn from a normal population. The method of moments has been studied as being the most useful. The calculation of percentage limits has been investigated with a view to using these findings in control chart analysis. In Part II the calculation of the standard deviation from the range has been considered and this estimate compared with those obtained by more rigorous methods. The control chart for sample ranges has been constructed together with that for sample deviations. The similarity between these two charts has been the basis for the conclusions of Part II.

determine the simultaneous distribution of any E-order stationary. The development given by wilks has been

Dith z₁, z₂, ..., z_n, a sample of size m from a cop

z, z, ..., z, arrenged in escending order or magnitude,

S.B.Wilks, Mathematical Statistics (Princetons

Princeton University Press, 19441, pp. 87-90

has been made to collect all svailable material dealing with the distribution of range in general, and in perticular with the distribution of range when the samples are drawn from a normal population. The method of moments has been studied as being the most useful. The establation of percentage limits has been investigated with a view to using these findings in control onart shalpsis. In Part 11 the calculation of the standard deviation from the range has been considered and this estimate compared with those obtained by more rigorous methods. The control obert for sample ranges has been constructed together with that for sample deviations. The similarity between these two charts has been the basis

PART I

THEORY OF RANGE

If a statistical measure is to be a useful tool, its nature must be thoroughly understood. The character of its distribution must be known. The distribution of range may be approached from two different standpoints. It may be developed from the distribution of K-order statistics or the moments of range may be computed with a view to finding the best-fitting curve. Both methods have been here considered.

I. SOME SPECIAL DISTRIBUTIONS

<u>Distribution of any K-order statistics</u>. Before considering the distribution of the range, it is necessary to determine the simultaneous distribution of any K-order statistics. The development given by Wilks has been followed in this part of the discussion.

With \mathbf{x}_1 , \mathbf{x}_2 , ..., \mathbf{x}_n , a sample of size n from a population with probability element $\mathbf{f}(\mathbf{x})$ dx, and with \mathbf{x}_1 , \mathbf{x}_2 , ..., \mathbf{x}_n , arranged in ascending order of magnitude, he let \mathbf{r}_1 , \mathbf{r}_2 , ..., \mathbf{r}_k , be k integers such that

S.S.Wilks, <u>Mathematical Statistics</u> (Princeton: Princeton University Press, 1944), pp. 89-90.

THEORY OF BANCE

If a statistical measure is to be a useful tool, its nature must be thoroughly understood. The character of its distribution must be known. The distribution of range may be approached from two different standpoints. It may be developed from the distribution of K-order statistics or the moments of range may be computed with a view to flading the best-fitting curve. Both methods have been here considered.

I. SOME SIRCIAL DISTRIBUTIONS

Distribution of any L-order statistics. Before considering the distribution of the range, it is necessary to determine the simultaneous distribution of any K-order statistics. The development given by wilks has been rollowed in this part of the discussion.

With x_1, x_2, \ldots, x_n , a sample of size n from a population with probability element f(x) dx, and with x_1, x_2, \ldots, x_n , arranged in ascending order of magnitude, he let r_1, r_2, \ldots, r_k , be k integers such that

S.S.Wilks, Mothematical Statistics (Frinceton: Trinceton University Tess, 1944), pp. 89-90.

 $1 \leqslant r_1 < r_2 < \ldots < r_k \leqslant n$. His problem was to find the probability element of $x_{r_1}, x_{r_2}, \ldots, x_{r_k}$, or

$$f(x_{r_1}, x_{r_2}, \ldots, x_{r_k}) dx_{r_1} dx_{r_2} \ldots dx_{r_k}$$

He considered the sample to be covered by 2k+1 intervals in such a way that each of the k elements x_{r_1} , x_{r_2} , ..., x_{r_k} has its own interval, the other k+1 intervals covering the remaining n-k elements of the sample, no two intervals overlapping. These intervals,

$$(-\infty, x_{r_1}), (x_{r_1}, x_{r_1} + dx_{r_1}), (x_{r_1} + dx_{r_1}, x_{r_2}), \dots,$$

$$(x_{r_k} + dx_{r_k}, \infty)$$
:

with

$$\int_{I_{i}} f(x) dx = q_{i} : \qquad (i = 1, 2, ..., 2k + 1)$$

$$\int_{i=1}^{2k+1} q_{i} = 1$$

The problem resolved itself into finding the probability (to terms of order $\mathrm{dx}_{r_1}, \, \mathrm{dx}_{r_2}, \, \ldots \, \mathrm{dx}_{r_k}$) that if a sample of n elements is drawn from a multinomial population with classes $I_1, \, I_2, \, \ldots, \, I_{2k+1}, \,$ then r_1 - 1 will fall in I_1 , 1 element in $I_2, \, r_2$ - r_1 -1 elements in I_3 , 1 element in $I_4, \, \ldots, \, n$ - r_k in I_{2k+1} . From the multinomial law it follows that the probability of such a partition is

 $1 \leqslant r_1 < r_2 < \dots < r_k \leqslant n$. His problem was to find the probability element of $x_{r_1}, x_{r_2}, \dots, x_{r_k},$ or $f(x_{r_1}, x_{r_2}, \dots, x_{r_k})$ dxr₁. dxr₂...dxr_k.

He considered the sample to be covered by 2k+1 intervals in such a way that each of the k elements x_{r_1} , x_{r_2} , ..., x_{r_k} has its own interval, the other k+1 intervals covering the remaining n - k elements of the sample, no two intervals overlapping. These intervals,

II, I2, 13, ..., IZK4], WETE

(- co , xr1), (xr1, xr1 + dxr1), (xr1 + dxr1, xr2), (rx , co -)

(xr + dxr =):

with $\begin{cases} f(x) \ dx = q_1 : \\ f(x) \ dx = 1 \end{cases}$

The problem resolved itself into finding the probability (to terms of order dx_1 , dx_2 , ... dx_k) that if a sample of n élements is drawn from a multinomial population with classes I₁, i₂, ..., I_{2k+1}, then r₁ - 1 will fall in I₁, 1 element in i₂, r₂-r₁-1 elements in I₃, 1 element in I_k, i_i, n-r_k in I_{2k+1}. From the multinomial law it rollows that the probability of such a portition is

$$q_1^{r_1-1}$$
 q_2^{1} $q_3^{r_2-r_1-1}$ q_4^{1} ... $q_{2k+1}^{n-r_k}$,

which becomes, when the values of qi are substituted,

$$\left[\int_{-\infty}^{x_{r_1}} f(x) dx \right]^{r_1-1} \left[\int_{x_{r_1}}^{x_{r_1}} f(x) dx \right] \left[\int_{x_{r_1}+dx_{r_1}}^{x_{r_2}} f(x) dx \right]^{r_2-r_1-1}$$

$$\begin{bmatrix} x_{r_2} + dx_{r_2} \\ x_{r_2} & f(x) dx \end{bmatrix} \cdots \begin{bmatrix} x_{r_k} + dx_{r_k} \end{bmatrix}^{n-r_k}$$

Now, to within terms of order dxri,

$$x_{r_{i}} + dx_{r_{i}}$$

$$\int_{x_{r_{i}}}^{x_{r_{i}}} f(x) dx = f(x_{r_{i}}) dx_{r_{i}} \text{ and } \int_{x_{r_{i}} + dx_{r_{i}}}^{x_{r_{i}} + 1} f(x) dx$$

$$= x_{r_{i+1}}$$

$$\int_{x_{r_i}}^{x_{r_i}} f(x) dx.$$

Hence, $f(x_{r_1}, x_{r_2}, \ldots, x_{r_k})$ $dx_{r_1} dx_{r_2} \ldots dx_{r_k}$

which becomes, when the values of q are substituted,

Now, to within terms of order dry,

$$\int_{\mathbb{X}_{\Gamma_1}} f(x) dx = f(x_{\Gamma_1}) dx_{\Gamma_1} \text{ and } \int_{\mathbb{X}_{\Gamma_1}} f(x) dx$$

$$\int_{\mathbb{R}^{n}} f(x) dx.$$

Hence, fix1, x2, ..., x2, dx1. dx2. ... dxx

$$= \frac{n!}{(r_1-1)! (r_2-r_1-1)! \dots (n-r_k)!} \left[\int_{\infty}^{r_1} f(x) dx \right]^{r_1-1} \left[\int_{x_{r_1}}^{x_{r_2}} f(x) dx$$

$$f(x_{r_1}) dx_{r_1} f(x_{r_2}) dx_{r_2} \dots f(x_{r_k}) dx_{r_k}$$

Distribution of the largest and smallest values in a sample. To apply the same technique to the joint distribution of the largest and smallest values of x in a sample, 5 intervals have been taken, i.e., 2K+1=5. Since $r_1=1$ and $r_k=r_2=n$, it is necessary to consider the probability of obtaining 0 elements in I_1 , 1 element in I_2 , n-2 elements in I_3 , 1 element in I_4 , and 0 elements in I_5 . From the preceding discussion, it follows that:

$$f(x_{1}, x_{n}) dx_{1} dx_{n} = \frac{n!}{0!1!(n-2)!1!0!}$$

$$\left[\int_{-\infty}^{x_{1}} f(x)dx\right]^{0} \left[\int_{x_{1}}^{x_{1}} f(x)dx\right] \left[\int_{x_{1}+dx_{1}}^{x_{1}} f(x)dx\right]^{n-2}$$

$$\begin{bmatrix} x_n + dx_n \\ x_n \end{bmatrix} f(x) dx \begin{bmatrix} x_n + dx_n \end{bmatrix}^{\circ}$$

Distribution of the largest and smallest values in a sample. To apply the same technique to the joint distribution of the largest and smallest values of x in a sample, 5 intervals have been taken, i.e., 2X+1=5. Since $r_1=1$ and $r_K=r_2=n$, it is necessary to consider the probability of obtaining 0 elements in l_1 , 1 elements in l_2 , n-2 elements in l_3 , 1 element in l_4 , and 0 elements in l_5 . From the preceding discussion, it follows that:

$$f(x_1, x_n) = \frac{n!}{0!1!(n-2)!1!0!}$$

$$\begin{bmatrix} x_1 & f(x) = x_1 \\ f(x) & f(x) \end{bmatrix} \begin{bmatrix} x_1 & f(x) = x_1 \\ f(x) & f(x) = x_1 \\ f(x) & f(x) = x_1 \end{bmatrix}$$

$$\begin{bmatrix} x_n + dx_n \\ 2n \end{bmatrix} \begin{bmatrix} x_n + dx_n \\ 2n \end{bmatrix} \begin{bmatrix} x_n + dx_n \end{bmatrix}.$$

The making of the substitutions, as before, gives

$$f(x_1, x_n) dx_1 dx_n = n (n-1) \left[\int_{x_1}^{x_n} f(x) dx \right]^{n-2}$$

$$f(x_1) dx_1 f(x_n) dx_n$$
.

Distribution of range. Wilks 10 then obtained the distribution of the sample range by letting

$$x_n - x_1 = R$$

 $x_1 = S$

and integrating the resulting distribution with respect to S. He illustrated this method by means of the rectangular distribution.

$$f(x) = 1/r \qquad 0 < x < r$$
= 0, otherwise.

This problem is perfectly straightforward, but difficulty is encountered when the distribution is of a less simple nature.

II. MOMENTS OF RANGE

Tippett11 says that the distribution of range cannot

^{10 &}lt;u>Ibid.</u>, p. 92.

¹¹ L. H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika, XVII (1925) Parts 3 and 4, 368.

The making of the substitutions, as before, gives

$$f(x_1, x_n) dx_1 dx_n = n (n-1) \begin{bmatrix} x_n & f(x) dx \\ 1 & 1 \end{bmatrix}^{n-2}$$

f(x1) dx1 f(xn) dxn.

Distribution of range by letting

$$x_n - x_1 = R$$

and integrating the resulting distribution with respect to S. He illustrated this method by means of the rectangular distribution.

$$f(x) = 1/x \qquad \qquad 0 < x < x$$

= 0, otherwise.

This problem is perfectly straightforward, but difficulty is encountered when the distribution is of a less simple nature.

II. MOLENTS OF TAMES

Tippett11 says that the distribution of range cannot

^{10 1}bid., p. 92.

¹¹ L. H. C. Tippett, "On the Extrane individuals and the Hange of Samples Taken from a Normal Population," Bio-metrika, XVII (1925) Parts 3 and 4, 368.

be written in any useful form, that the problem has been to find the first four moments so that an appropriate Pearson curve can be fitted, adding that such curves fit actual data sufficiently well to establish the adequacy of the method for practical purposes.

<u>Deriving the equations</u>. The first method of finding the mean range considered by Tippett involves the use of Karl Pearson's 12 expression for the mean difference between the p^{th} and (p+1)th individual, the expression having been obtained in the following manner:

The frequency distribution was represented by $y=N\varphi(x)$, with no hypothesis as to the nature of the distribution. N was the number of individuals; A, the area to the left of any ordinate y at a character value x; N - A, the area to the right. Then if $\phi=A/N$

$$d = \int_{-\infty}^{x} \phi(x) dx$$
.

The chance of any random individual having a character value less than x is A/N = A and the chance of having a character value greater than x is (N - A)/N = 1 - A.

 x_p corresponded to the pth individual's character, and x_{p+1} to the next or (p+1)th individual's character. The

¹² Karl Pearson, "Note on Francis Galton's Problem," Biometrika, I (1902), 391-92.

be written in any useful form, that the problem has been to find the first four moments so that an appropriate Pearson curve can be fitted, adding that such curves fit actual data sufficiently well to establish the adequacy of the method for practical purposes.

Deriving the equations. The first method of finding the mean range considered by Tippett involves the use of Marl Pearson's 2 expression for the mean difference between the pth and (p+1)th individuel, the expression having been obtained in the following manner:

The frequency distribution was represented by $y = N \phi(x)$, with no hypothesis as to the nature of the distribution. N was the number of individuals; A, the area to the left of any ordinate y at a character value x; N - A, the area to the right. Then if d = A/N d = A/N

The chance of any random individual having a character value less than x is A/N = d and the chance of having a character value greater than x is (N-A)/N = 1 - d.

 x_p corresponded to the p^{th} individual's character, the next or (p+1)th individual's character. The

¹² Karl Pearson, "Note on Francis Galton's Problem," Biometrika, 'I (1902), 391-92.

problem was to find the mean value $x_p - x_{p+1}$, there being p-1 individuals to the right of y_p and n-p-1 to the left of y_{p+1} in a sample of size n. The chance of an individual falling at x_p is given by $y_p dx_p/N$, and one at x_{p+1} , by $y_{p+1} dx_{p+1}/N$; the chance of an individual to the left of y_{p+1} is A_{p+1}/N and to the right of Y_p is $(N-A_p)/N$. The total chance of this combination is given by

$$\frac{y_p \, dx_p}{N} \cdot \frac{Y_{p+1} \, dx_{p+1}}{N} \cdot \left[\begin{array}{c} A_{p+1} \\ \end{array} \right]^{n-p-1} \left[\begin{array}{c} N - A_p \\ \end{array} \right] \cdot$$

But the two individuals can be permuted in n!/ (n-p-1)! (p-1)! ways. To get the average Pearson multiplied the chance thus obtained by the corresponding $x_p - x_{p+1}$, and integrated from $x_{p+1} = -\infty$ to x_p and for $x_p = -\infty$ to ∞ , writing x' for x_{p+1} , x for x_p , A for A_{p+1}/N , A for A_p/N , A for A

$$X_{p} = \frac{n!}{(n-p-1)!(p-1)!} - \int_{\infty}^{\infty} dx \int_{-\infty}^{x} dx' \quad y_{0}. \quad y_{0}' \quad \alpha' \quad (1-\alpha)^{p-1}$$

$$(x-x')$$

where

$$\frac{d\mathbf{d'}}{d\mathbf{x'}} = y_0' , \quad \frac{d\mathbf{d}}{d\mathbf{x}} = y_0 .$$

II

record and the control of the contro

and some one to the control of the c

023,60

Now the x' integral was considered.

$$I = \int_{-\infty}^{x} y_{0}' d^{n-p-1} (x-x') dx'$$

$$= \int_{-\infty}^{x} a^{n-p-1} (x-x') da'.$$

Integration by parts gave

$$\left[\frac{d^{n-p}}{n-p} \quad (x-x^{*})\right]_{\infty}^{x} + \int_{-\infty}^{x} \frac{d^{n-p}}{n-p} dx^{*}$$

or between the limits

$$\frac{1}{n-p} \int_{-\infty}^{x} d' \frac{n-p}{dx'} = \frac{1}{n-p} U, \text{ say.}$$

Thus

$$X_{p} = \frac{n!}{(n-p)! (p-1)!} \int_{-\infty}^{\infty} y_{0} U (1-\alpha)^{p-1} dx$$

$$= \frac{n!}{(n-p)! (p-1)!} \int_{-\infty}^{\infty} U (1-\alpha)^{p-1} d\alpha$$

$$= \frac{n!}{(n-p)! p!} \left\{ \left[-U (1-\alpha)^{p} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{dU}{dx} (1-\alpha)^{p} dx \right\}$$

.bereblence asw fergetal 'x ent woll

$$I = \int_{-\infty}^{\infty} e^{x} y_{0}' \quad d^{-p-1} \quad (x-x') \quad dx'$$

$$= \int_{-\infty}^{\infty} e^{x} e^{-p-1} \quad (x-x') \quad d\alpha'.$$

integration by parts gave

$$\begin{bmatrix} a_{-p} & (x-x) \end{bmatrix} + \begin{bmatrix} x & a_{-p} & ax \\ a_{-p} & ax \end{bmatrix}$$

or between the limits

shill

$$= \frac{1}{2} \left\{ \frac{(n-p) \cdot (p-1) \cdot (p-1)$$

or taking the limits and substituting $\frac{dU}{dx}$

$$x_p = \frac{n!}{(n-p)! p!} \int_{-\infty}^{\infty} a^{n-p} (1-\alpha)^p dx.$$

This then was the formula used by Tippett. The sum of these mean differences for all values of p from 1 to n-1 gives the mean range, $\overline{\mathbf{w}}$.

$$\overline{w} = \int_{-\infty}^{\infty} \left[1 - (1-\alpha)^n - \alpha^n \right] dx.$$

Tippett¹³ also gave a second method of finding the mean range. Figure 1 represents the curve of the distribution of the original population, $y = \varphi(x)$, and as before

$$a_p = \int_{-\infty}^{x_p} \phi(x) dx$$
, where $\int_{-\infty}^{\infty} \phi(x) dx = 1$.

If \mathbf{x}_1 is the character of the first individual, and \mathbf{x}_n that of the last in a sample of size n, then, on the assumption that the original population is infinite, the chance that there is one individual at \mathbf{x}_1 , one at \mathbf{x}_n , and n-2 between is given by

L. H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," 368-70.

Uh gnijutivedus bns slimil sut gnikst to

$$x_{\bar{p}} = \frac{n!}{(n-\bar{p})! \, \bar{p}!} \int_{-\infty}^{\infty} d^{n-\bar{p}} \, (1-\alpha)^{\bar{p}} \, dx.$$

This then was the formula used by Tippett. The sum of these mesh differences for all values of p from 1 to n-1 gives the mean range, w.

$$\overline{\mathbf{w}} = \int_{-\infty}^{\infty} \left[1 - (1-\omega)^{\Omega} - \omega^{\Omega} \right] d\mathbf{x}.$$

Tippett also gave a second method of finding the mean range. Figure 1 represents the curve of the distribution of the original population, $y = \omega(x)$, and as before

$$d_{\mathcal{G}} = \int_{-\infty}^{\infty} dx (x) dx$$
, where $\int_{-\infty}^{\infty} dx = 1$.

If \mathbf{x}_1 is the character of the first individuel, and \mathbf{x}_n that of the lest in a sample of size n, then, on the assumption that the original population is infinite, the chance that there is one individual at \mathbf{x}_1 , one at \mathbf{x}_n , and \mathbf{n} -2 between is given by

I. H. C. Tippett, "On the Extreme Individuels and the Range of Samples Texen from a North Population," 368-70.

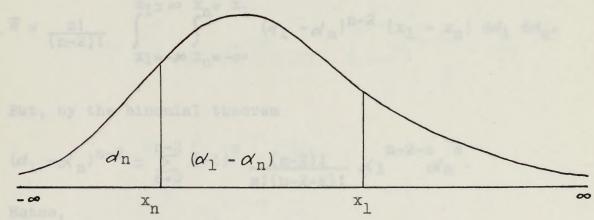


FIGURE 1 CURVE OF DISTRIBUTION, $y = \phi(x)$

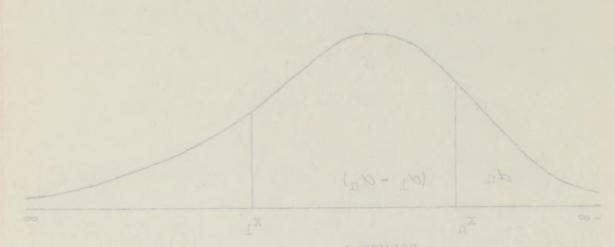


FIGURE 1 CURVE OF DISTRIBUTION, y = O(x)

$$\frac{n!}{(n-2)!}$$
 $(d_1 - d_n)^{n-2} dd_1 dd_n$.

This expression is equivalent to that of Wilks, except for the fact that Wilks arranged his individuals in ascending order of magnitude; Tippett, in descending. The notion of expected value leads to the defining of the mean range as

$$\overline{w} = \frac{n!}{(n-2)!} \int_{x_1=-\infty}^{x_1=\infty} x_n = x_1 (d_1 - d_n)^{n-2} (x_1 - x_n) dd_1 dd_n.$$

But, by the binomial theorem

$$(d_1 - d_n)^{n-2} = \sum_{s=0}^{n-2} (-1)^s \frac{(n-2)!}{s!(n-2-s)!} d_1^{n-2-s} d_n^s$$

Hence,

$$\overline{w} = n! \sum_{s=0}^{n-2} \frac{(-1)^s}{s!(n-2-s)!} \int_{x_1=-\infty}^{x_1=\infty} d_1^{n-2-s} dd_1 \int_{x_n=-\infty}^{x_n=x_1} d_n^{s}(x_1-x_n) dd_n.$$

$$\int_{x_{n}=-\infty}^{x_{n}=x_{1}} d_{n}^{s}(x_{1}-x_{n}) dd_{n} = \left[\frac{(x_{1}-x_{n})a_{n}^{s+1}}{s+1} \right]_{x_{n}=-\infty}^{x_{n}=x_{1}}$$

+
$$\frac{1}{s+1}$$
 $\int_{x_n=-\infty}^{x_n=x_1} \alpha_n^{s+1} dx_n$.

This expression is equivalent to that of Wilks, except for the fact that Wilks arranged his individuals in ascending order of magnitude; Tippett, in descending. The notion of expected value leads to the defining of the mean range as

But, by the blocklal theorem

$$(d_1 - d_n)^{n-2} = \sum_{s=0}^{n-2} (-1)^s \frac{(n-2)!}{s!(n-2-s)!} d_1 d_n$$

Hence,

$$\overline{w} = n! \sum_{n=0}^{n-2} \frac{(-1)^n}{s!(n-2-s)!} \int_{-\infty}^{x_1 = \infty} \int_{-\infty}^{x_2 = x_1} du^{-2+s} \, dd_1 \int_{-\infty}^{x_2 = x_1} du^{-s}(x_1 - x_1) \, dd_2.$$

$$\sum_{x_{n}=-\infty}^{\infty} d_{n}^{2}(x_{1}-x_{n}) dd_{n} = \underbrace{(x_{1}-x_{n})d_{n}^{2}+1}_{x_{1}} dd_{n}^{2} = \underbrace{(x_{1}-x_{n})d_{n}^{2}+1}_{x_{1}}$$

The term in brackets vanishes at both limits, since

$$x_1 = x_n = 0$$
, when $x_n = x_1$,

 $a_n^{s+1} = \int_{-\infty}^{x_n} \phi(x) dx$ s+1 = 0, when $x = -\infty$,

so that, by substitution,

$$\overline{w} = n! \sum_{s=0}^{n-2} \frac{(-1)^s}{(s+1)! (n-2-s)!} \int_{x_1=-\infty}^{x_1=\infty} q^{n-2-s} U dq_1,$$

where

$$U = \int_{-\infty}^{x_1} \alpha_n^{s+1} dx_n.$$

If
$$\theta = \int_{a_1}^1 a_1^{n-2-s} da_1$$

$$= \frac{1-\alpha_1^{n-1-s}}{n-1-s},$$

then

$$\frac{d\theta}{d\alpha_1} = -\alpha_1^{n-2-s}$$

The term in brackets vanishes at both limits, since

$$x_1 = x_1 = 0$$
, when $x_1 = x_1$,

and
$$d_n^{s+1} = \begin{bmatrix} x_n & \phi(x) & dx \end{bmatrix}^{s+1} = 0$$
, when $x = -\gamma$,

so that, by substitution,

where

If
$$\theta = \int_{c_{1}}^{b} c_{1}^{n-2-a} dc_{1}$$

nend

and
$$\overline{w} = -n! \sum_{s=0}^{n-2} \frac{(-1)^s}{(s+1)! (n-2-s)!} \int_{x_1=-\infty}^{x_1=\infty} U \frac{d\theta}{d\alpha_i} d\alpha_1.$$

Integration by parts gives

$$\overline{w} = n! \sum_{s=0}^{n-2} \frac{(-1)^{s}}{(s+1)! (n-2-s)!} \left\{ -\left[u e \right] \begin{array}{c} x_{1} = \infty \\ x_{1} = -\infty \end{array} \right.$$

$$+ \int_{-\infty}^{\infty} e \frac{du}{dx_{1}} dx_{1} \right\}.$$

Since the term in brackets vanishes and

$$\frac{dU}{dx_1} = \frac{d}{dx_1} \int_{-\infty}^{x_1} dx_n = \frac{d}{dx_1} \int_{-\infty}^{x_1} \left[\int_{-\infty}^{x_n} \mathcal{O}(x) dx \right]_{0}^{x_n}$$

$$=\alpha_1^{s+1}$$

$$\overline{w} = \sum_{s=0}^{n-2} (-1)^s \frac{n!}{(s+1)! (n-1-s)!} \int_{-\infty}^{\infty} (1 - \alpha_1^{n-1-s}) \alpha_1^{s+1} dx,$$

which leads to Tippett's previous result

$$\overline{w} = \int_{-\infty}^{\infty} \left[1 - (1 - \alpha)^n - \alpha^n\right] dx.$$

and
$$\overline{w} = -n$$
: $\frac{1}{8} = \frac{(-1)^{\frac{1}{2}}}{(n-2-n)!} = \frac{1}{3} = \frac{1}{3}$ and $\frac{1}{3} = \frac{1}{3} = \frac{1}{3}$ and $\frac{1}{3} = \frac{1}{3} =$

integration by parts gives

$$\begin{cases} -1 & \frac{1}{1} & \frac{1}{1}$$

Since the term in brackets vanishes and

$$\frac{dx_1}{dx_2} = \frac{dx_1}{dx_1} = \frac{dx_1}{dx_2} = \frac{dx_2}{dx_2} = \frac{dx_2}{dx_2$$

$$\overline{W} = \sum_{n=0}^{\infty} (-1)^n \frac{(n+1)! (n+1-n)!}{(n+1-n)! (n+1-n)!} - \sum_{n=0}^{\infty} (1-n)^n \frac{1}{n} = 0$$

which leads to Tippett's previous result

$$\overline{w} = \int_{-\infty}^{\infty} \left[1 - (1 - \alpha)^n - \alpha^n \right] dx.$$

The higher moments may be obtained in a similar manner so that for even values of n

$$\mu_{m} = m (m-1)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{x_{1}} \left[1 - d_{1}^{n} - (1 - d_{n})^{n} + (1 - d_{n})^{n}\right]$$

$$(x_1 - x_n - \overline{w})^{m-2}$$
 $dx_1 dx_n - (m-1) (-\overline{w})^m$

On putting m = 2,

$$\mu_2 = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{x_1} \left[1 - \alpha_1^n - (1 - \alpha_n)^n\right]$$

$$+ (_{1} - \alpha_{n})^{n}] dx_{1} dx_{n} - \overline{w}^{2}$$
.

Constructing the tables. From the equation for the mean range when the samples are from a normal population Tippett found a framework of values by direct computation, using quadratures. This he filled in by interpolation using first Lagrangian Formulae; and lastly, a difference formula. The result was his table for the mean range for a normal distribution for samples of size n from 2 to 1000. The values are for a population having unit standard deviation,

^{14&}lt;u>Tbid.</u>, pp. 371-73.

The higher moments may be obtained in a similar manner o that for even values of n

$$\mathcal{L}_{m} = m \ (m-1) \ \mathcal{L}_{m} = \frac{1}{2} \ \left[1 - q_{1} - (1 - q_{1})_{1} + (1 - q_{2})_{1} \right]$$

$$(x_1 - x_1 - \overline{w})^{m-2} dx_1 dx_1 - (m-1) (-\overline{w})^{m}$$

On putting m = 2, m = 2,

Constructing the tables. From the equation for the

dead range when the samples are from a normal population, Tippett tound a fragework of values by direct computation, using quadratures. This he filled in by interpolation using first lagrangian Formulae; and lastly, a difference formula. The result was his table for the mean range for a normal distribution for samples of size n from 2 to 1000. The values are for a population having unit standard deviation,

¹⁴ Ibid., pp. 371-73.

so that to obtain the absolute range in any given case, the tabled value must be multiplied by the actual value of the standard deviation. Figure 2 illustrates Tippett's results graphically.

A similar method was used in the case of the second moment and the standard deviation. The framework was constructed by substitution in the formula. The process involving cubature was very laborious. His results have been shown in Figure 3.

Much time was spent in trying to evaluate the third and fourth moments by this same method, but many difficulties were encountered and the results obtained were irregular. One cause of difficulty was the fact that the equation consists of two nearly equal parts, one of which must be subtracted from the other, so that the computations must be very accurate if the difference is to be relied upon. Consequently, Tippett resorted to a method of obtaining μ_3 and μ_4 from the separate distributions of the first and last individuals. He started with the following general formulas:

$$\overline{w} = \overline{u} - \overline{v}$$
 $\mu_{3w}^{2w} = 2\mu_{2u}(1-r)$
 $\mu_{3w}^{2} = 2\mu_{3u}^{2u} + 6p_{12}$
 $\mu_{4w}^{2} = 2\mu_{4u}^{2u} - 8p_{13} + 6p_{22}$

where the p's are certain product moment coefficients.

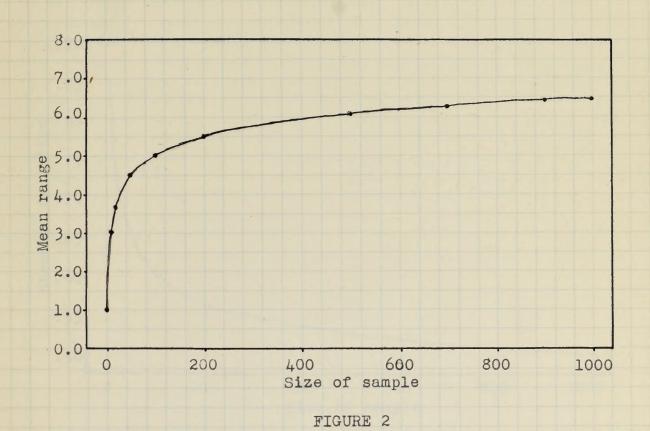
so that to obtain the absolute range in any given case, the tabled value must be multiplied by the actual value of the standard deviation. Figure 2 lilustrates Tippett's results graphically.

A similar method was used in the case of the second moment and the standard deviation. The framework was constructed by substitution in the formula. The process involving oubsture was very laborious. His results have been shown in Figure 3.

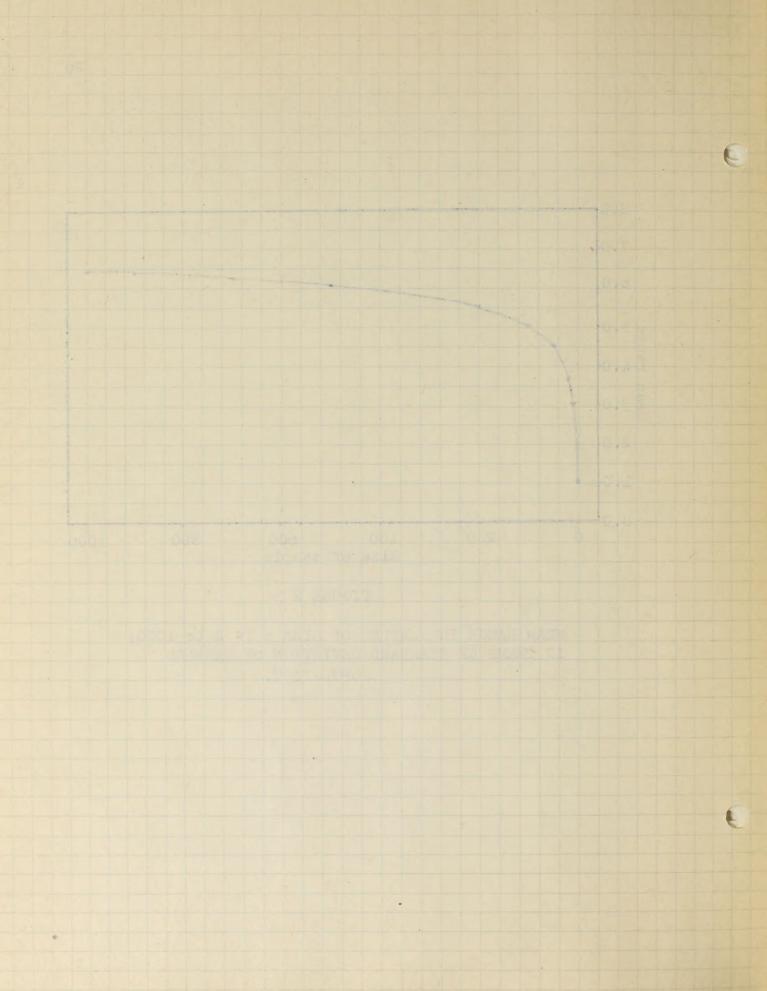
And fourth moments by this same method, but many difficulties were encountered and the results obtained were irregular.

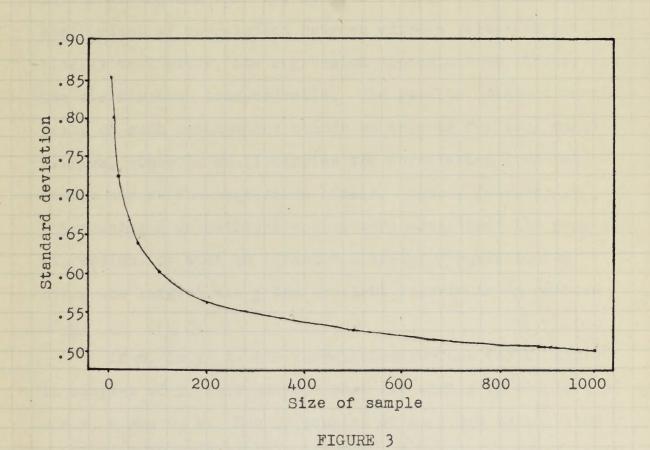
One cause of difficulty was the fact that the equation consists of two mearly equal parts, one of which must be subtracted from the other, so that the computations must be very accurate if the difference is to be reited upon. Consequently, Tippett resorted to a method of obtaining \$\mu_1\$ and \$\mu_4\$ from the separate distributions of the first and last individuals. He started with the following general formulas:

where the p's are certain product moment coefficients.

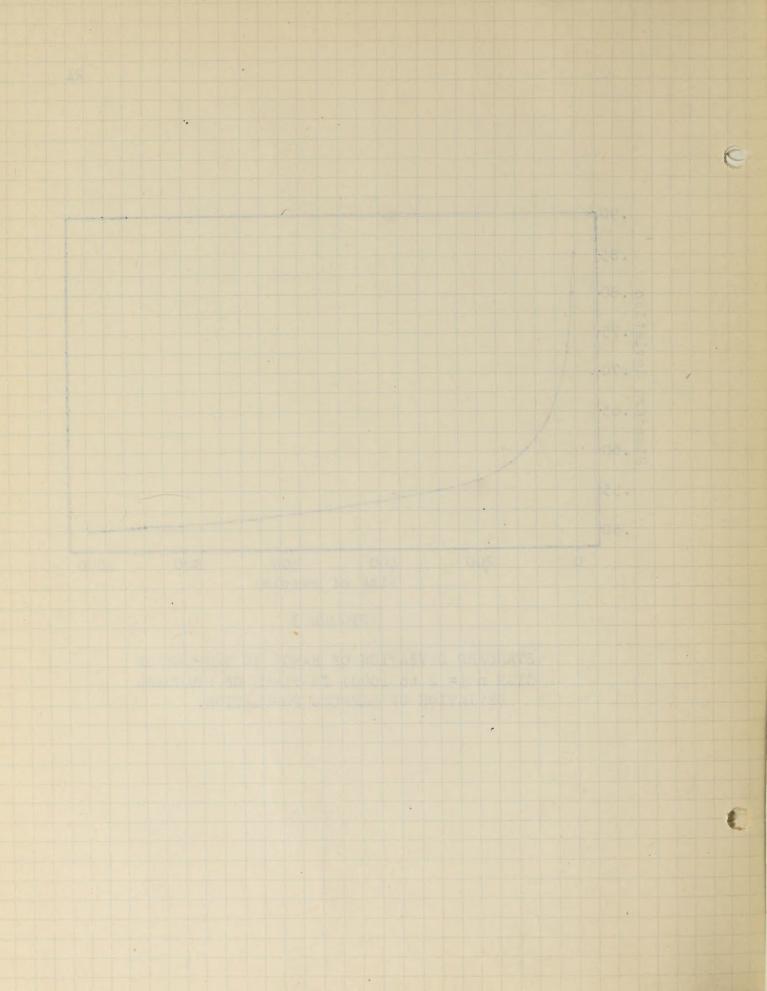


MEAN RANGE IN SAMPLES OF SIZE n (= 2 to 1000)
IN TERMS OF STANDARD DEVIATION OF SAMPLED
POPULATION.





STANDARD DEVIATION OF RANGE IN SAMPLES OF SIZE n (= 2 to 1000) IN TERMS OF STANDARD DEVIATION OF SAMPLED POPULATION.



E. S. Pearson, however, called attention to the fact that Tippett, in obtaining the moments, simplified these formulas on the assumption that for samples of size 60 or more the coefficient of correlation between extreme values is practically negligible, the regression approximately linear, and the distribution homoscedastic. The results obtained on the basis of such assumptions cannot be trusted for very small samples, since in small samples the correlation does not vanish nor is the regression linear. Hence, Pearson, by introducing a geometrical conception, considered the general expression for what he termed "...the $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$ surface, " $\underline{\mathbf{u}}$ being the largest and $\underline{\mathbf{v}}$ the smallest individual in samples of $\underline{\mathbf{n}}$.

If ω_n (u,v) is the correlation surface for u, and v in samples of n, this surface must lie wholly to the left of u = v, since u > v. Now if samples of n+1 are taken, the surface ω_n (U,V) differs from ω_n (u,v), since x, the n+1 th individual, may be such that u > x > v, or u > v > x, or x > u > v. If the sampling is from a normal population with unit standard deviation, the distribution of this single individual is

E. S. Pearson, "A Further Note on the Distribution of Range in Samples Taken from a Normal Population," Biometrika, XVII (1926), 173-93.

This are assembled in the standard of the standard of the standard on the standard of the stan

in and the or is, the state of the wholes in 1972 of the analysis of the state of t

of mean a calculation of the contraction, " . The contraction of the calculation of the c

$$\frac{1}{\sqrt{2\pi}} \quad e^{-\frac{1}{2}x^2}.$$

Pearson considered the frequency distribution of x in each case and by summing the three contributions corresponding to every value of x possible he obtained the complete frequency surface \mathscr{Q}_{n+1} (U, V):

$$\mathcal{Q}_{n+1}\left(\mathbf{U},\,\mathbf{V}\right) = \mathcal{Q}_{n}\left(\mathbf{U},\,\mathbf{V}\right) \int_{\mathbf{V}}^{\mathbf{U}} \frac{e^{-\frac{1}{2}\mathbf{x}^{2}}}{\sqrt{2\pi}} \,\mathrm{d}\mathbf{x} + \frac{e^{-\frac{1}{2}\mathbf{V}^{2}}}{\sqrt{2\pi}} \int_{\mathbf{V}}^{\mathbf{U}} \mathcal{Q}_{n}\left(\mathbf{U},\mathbf{v}\right) \,\mathrm{d}\mathbf{v}$$

$$+\frac{e^{-\frac{1}{2}V^2}}{\sqrt{2\pi}}\int_{V}^{U} \mathcal{Q}_{n}(u,V) du.$$

This provided a reduction formula for obtaining the correlation surface of the extreme individuals in samples of n in terms of the equations of surfaces for smaller samples. By further substitutions this formula was reduced to

$$z = \frac{n (n-1)}{2\pi} e^{-\frac{1}{2}(u^2 + v^2)}$$
 $(A_u - A_v)^{n-2}$

where
$$A_t = \int_{-\infty}^{t} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx$$
.

This equation together with that for the frequency distribution of \underline{u} , given by

Pearson considered the frequency distribution of x in each case and by summing the three contributions corresponding to every value of x possible he obtained the com lete frequency surface of x (U. V):

$$\omega_{n+1}(u, v) = \omega_n(u, v) \int_{v}^{u} \frac{-\frac{1}{2}x^2}{\sqrt{2}v} dx_{+} \frac{-\frac{1}{2}v^2}{\sqrt{2}v} \int_{v}^{u} dx_{-} (v, v) dv$$

This provided a resultion formula for obtaining the correlation surface of the entreme individuals in samples of n in terms of the equations of surfaces for smaller samples. By further substitutions this formula was reduced to

$$S = n \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \right) = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \right) = \frac{1}{\sqrt{2}}$$

where
$$\Delta t = \int_{-\infty}^{t} \frac{e^{-2x^2}}{\sqrt{2\pi}}$$
 dx.

This equation together with that for the frequency distribution of u, given by

$$\mathcal{O}(u) = \frac{n}{\sqrt{2\pi}} \quad e^{-\frac{1}{2}u^2} \quad A_u \quad n-1$$

when the sample is drawn from a normal population, was used to find the moments and product moment coefficients needed for Tippett's general relations for the moments of range. The expressions involved cannot in general be integrated, but Pearson completed the solution for the cases of n=2, 3, 4, 5, and 6, by using integrals which he evaluated by quadrature. He found that his values for the mean ranges corresponded exactly with those obtained by Tippett. The values he obtained for $G_{\rm W}$, are shown in Figure 4, which represents on an enlarged scale the start of Tippett's curve illustrated in Figure 3.

Pearson further investigated what justification there was for assuming linear regression and homoscedasticity at n = 6 in the case of the constants β_1 , and β_2 . The differences between the results obtained from the general expression and those based on the preceding assumptions were too great to warrant the use of these assumptions. In fact, he came to the conclusion that, however great n may be, the regression can never be strictly linear, nor can theoretical homoscedasticity be obtained over the whole surface. Still, both assumptions may be justified in the region of significant frequency, and as n increases the coefficient of correlation tends to zero. Hence, for

$$\sigma(u) = \frac{u}{\sqrt{2\pi}} \quad e^{-\frac{1}{2}u^2} \quad u_u \stackrel{n-1}{=} 1$$

when the sample is drawn from a normal jopulation, was used to find the maments and product moment coefficients needed for Tippett's general relations for the moments of range. The expressions involved cannot in general be integrated, but Feerson completed the solution for the cases of n = 2, 3, 4, 5, and o, by using integrals which he evaluated by quadrature. He round that his values for the mean ranges corresponded exactly with those obtained by Tippett. The values he obtained for \(\sigma_w \), are shown in Figure 4, which represents on an enlarged scale the atert of Tippett's curve illustrated in figure 3.

Pearson further investigated what justification there was for essuring linear regression and homoscedasticity at n = 6 in the case of the constants 2, and 2. The differences between the results obtained from the general expression and those based on the preceding assumptions were too great to warrant the use of these assumptions. In fact, he came to the conclusion that, however great n may be, the regression can never be strictly linear, nor can theoretical homoscedasticity be obtained over the whole surface. Still, both assumptions may be justified in the region of significent frequency, and as n increases the region of significent frequency, and as n increases the

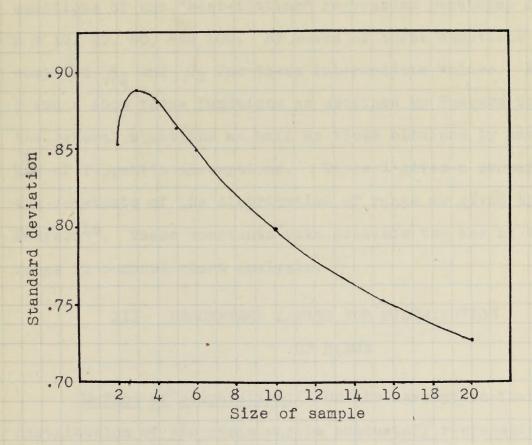
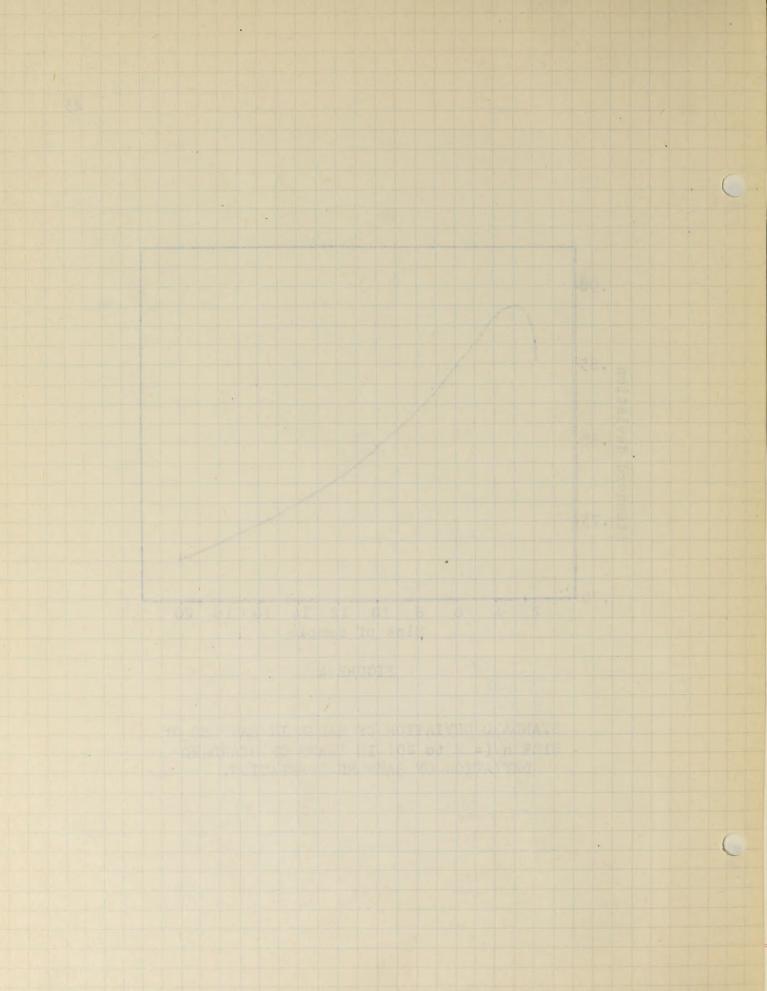


FIGURE 4

STANDARD DEVIATION OF RANGE IN SAMPLES OF SIZE n (= 2 to 20) IN TERMS OF STANDARD DEVIATION OF SAMPLED POPULATION.



samples of 60 or more serious error can hardly be involved in using Tippett's constants β_1 and β_2 . To bridge the gap between n=6 and n=60, Pearson determined the equations of the "best-fitting" regression parabolae for n=10, 20, 60, and 100. By means of these equations he computed β_1 and β_2 for these intermediate values. Figures 5 and 6 show these functions as obtained by Pearson from the general equations as well as those obtained by making use of Tippett's assumptions. Table I gives a summary of the constants of the distribution of range as given by Pearson¹⁶. These constants make possible the use of the range in control chart analysis.

III. PERCENTAGE LIMITS FOR DISTRIBUTION OF RANGE

Method of computation. On the assumption that the distribution of the range may be adequately represented by Pearson curves with appropriate moment coefficients, E. S. Pearson obtained a framework by finding equations of Type I and Type VI curves, using appropriate frequency con-

^{16&}lt;sub>Ibid.</sub>, p. 192.

¹⁷E. S. Pearson, "The Percentage Limits for the Distribution of Range in Samples from a Normal Population," Biometrika, XXIV (1932), 404-7.

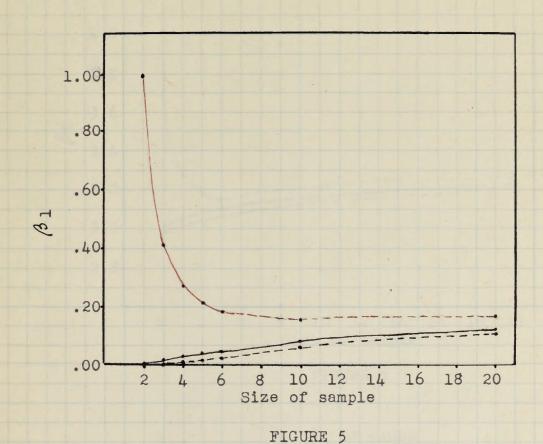
samples of 60 or more serious error osn hardly be involved in using Tippett's constants β_1 and β_2 . To bridge the gap between n=6 and n=60, Ferrson determined the equations of the "best-fitting" regression paracolae for n=10, 20, 60, and 100. By means of these equations he computed β_1 and β_2 for these intermediate values. Figures 5 and 6 show these functions as obtained by Festron from the central equations as obtained by Festron from use of Tippett's assumptions. Table I gives a summary of the constants of the distribution of range as siven by Festrond. These constants make possible the use of the range in control chart analysis.

III. PERCENTAGE LILITS FOR DISTRIBUTION OF RANGE

isting of computation. On the assumption that the distribution of the range may be adequately represented by Pearson curves with appropriate moment coefficients, E. S. Pearson 7 obtained a framework by finding equations of Type I and Type VI curves, using appropriate frequency con-

^{161919.} b. 192.

Distribution of Hange in Samples from a Mormal Population," Blometrika, XXIV (1932), 404-7.

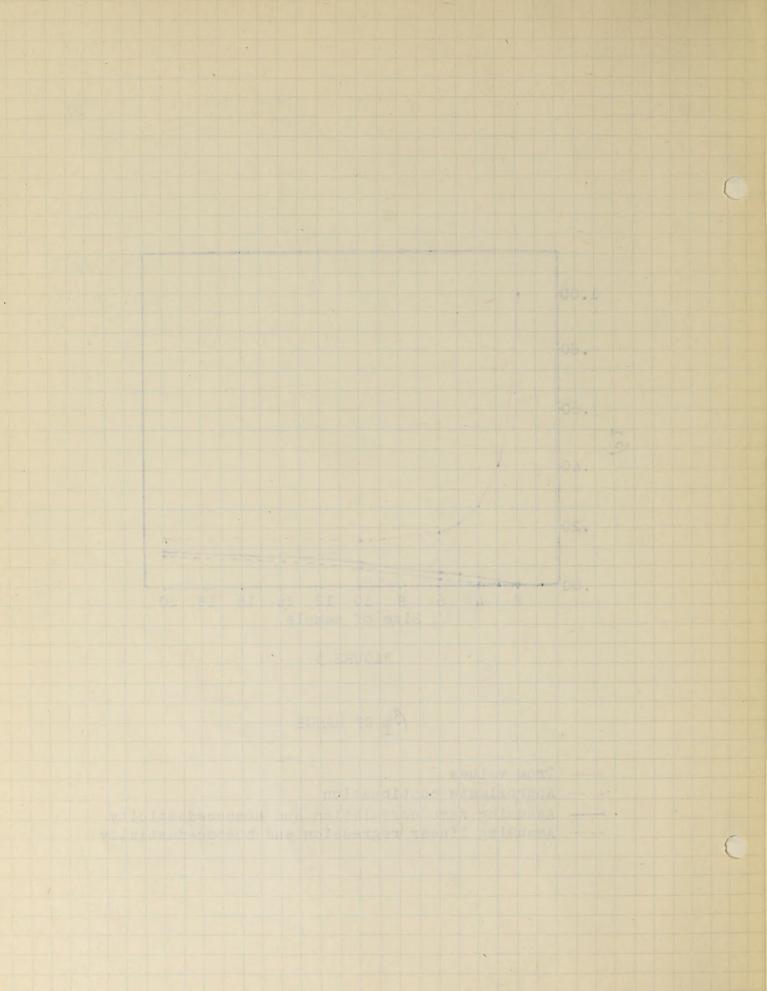


 β_1 OF RANGE

True values

-- Approximate continuation

--- Assuming zero correlation and homoscedasticity
---- Assuming linear regression and homoscedasticity



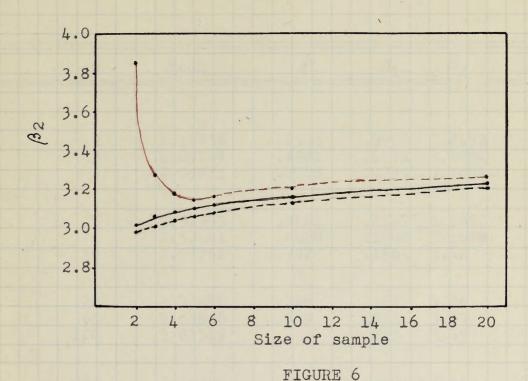


FIGURE O

 β_2 OF RANGE

True values

Approximate continuation

Assuming zero correlation and homoscedasticity
---- Assuming linear regression and homoscedasticity

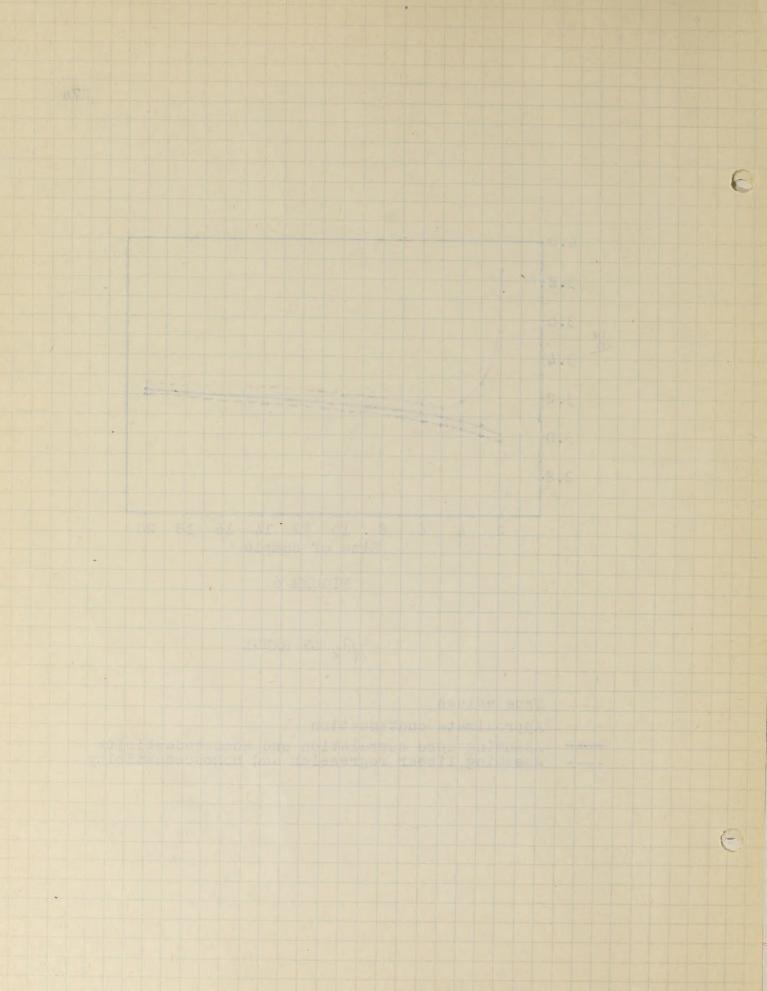


TABLE I
SUMMARY OF CONSTANTS OF DISTRIBUTION OF RANGE

n	Mean	6	Bl	B2
2 3 4	1.12838	.8525	.9906	3.8692
	1.69257	.8884	.4174	3.2864
	2.05875	.8798	.2735	3.1884
5	2.32593	.8641	.2167	3.1693
6	2.53441	.8480	.1892	3.1698
10	3.07751	.797	.156	3.22
20	3.73495	.729	.161	3.26
60	4.63856	.639	.201	3.35
100	5.01519	.605	.223	3.39
200	5.49209	.566	.247	3.44
500	6.07340	.524	.285	3.50
1000	6.48287	.497	.309	3.54

Sheli il dol'usirishe to emergent to year in

	1		4 (48)	
1 408 S. F.	1900 8 1 1971 1 1975 1	3585. 1883. 1879.	1.12338 1.89267 2.05575	
1.1893	Tols. Sear. of E.	1.168. 0848. .301.	2.7953 2.63441 3.67751	
98.6 8:5	161.	789 653 300,	20187. 00860.4 01280.1	
3.24 3.36 3.34	743. 688. 600.	4000. 400. 1104.	2.13203 0.18203 0.18287	

stants for samples of size

n = 3, 4, 6, 10, 20, 60, 100.

The first four curves were made to start at the point, range = w = o, and were given the correct mean (= \overline{w}) standard deviation (= G_W) and β_1 . For the other three curves the start was not fixed, but the first four theoretical moments used — \overline{w} , G_W , β_1 , and β_2 . The use of different methods is accounted for by the fact that when n is small the distribution of the range is abrupt at the lower end. Hence, it seemed advantageous to give the curves the correct start. On the other hand, as n increased, it seemed advisable to use the correct β_2 rather than the correct start.

The percentage limits computed were the upper and lower 0.5, 1, 5, and 10, thus giving the boundaries within which 99%, 98%, 90%, and 80% of the ranges would lie. The position of the ordinate at the upper and lower limits for each of the framework curves was found by quadrature and backward interpolation. For a given percentage limit, p, the value l_p (the position of the ordinate) changes with n, that is, with β_1 and β_2 or with the shape of the sampling curve. It was found that the change was not rapid so that it was possible to find by interpolation in the framework each of the eight values of l_p for

 $n = 3, 4, 5, \ldots, 29, 30, 35, 40, \ldots, 95, 100.$

stis to seignes tot sines

n = 3, 4, 6, 10, 20, 60, 100.

The first four ourves were given the correct mean $(=\overline{w})$ stendard range = w = o, and were given the correct mean $(=\overline{w})$ stendard deviation $(=\overline{G_w})$ and β_1 . For the other three curves the start was not fixed, but the first four theoretical moments used $-\overline{w}$, G_w , β_1 , and β_2 . The use of different methods is soccunited for by the fact that when n is small the distribution of the range is abrupt at the lower end. Hence, it seemed advantageous to give the curves the correct start. On the other hand, as n increased, it seemed advisable to use the correct β_2 rather than the correct start.

The percentage limits occurred were the upper and lower 0.5, 1, 5, and 10, thus giving the boundaries within which 95%, 95%, 90%, and 80% of the ranges would lie. The position of the ordinate at the upper and lower limits for sech of the framework durves was found by quadrature and backward interpolation. For a given percent selimit, p, the value in the position of the ordinate) changes with n, that is, with \$\beta_1\$ and \$\beta_2\$ or with the shape of the shapling curve. It was found that the change was not rapid so that to was possible to find by interpolation in the framework sech of the cint values of 1p for

n = 3, 4, 5,, 29, 30; 35, 40, 95, 100.

Since $l_p = (w_p - \overline{w})/\sigma_w$, where w_p represents the range value corresponding to any one of the ordinates, having calculated the l_p 's, it was only necessary to invert the formula to obtain w_p from

$$w_p = \overline{w} + 1_p \sigma_w$$
.

Pearson used values of \overline{w} as computed by Tippett, but $\textit{\textbf{G}}_{W}$ had only been calculated for

n = 2, 3, 4, 5, 6, 10, 20, 60, 100.

Three additional values were computed at

by the same process of cubature as that employed by Tippett. From this framework the intermediate values of \mathcal{T}_W were obtained and finally the values of \mathbb{I}_p given in Pearson's Table A. Since the values of \mathbb{W} and \mathcal{T}_W are for samples drawn from a normal population, this table gives the percentage limits for the distribution of range in samples from a normal population.

In a more recent article Pearson¹⁹ stated that no simple expression exists for the probability law f_n (w) of w, but he gave a table of computed values of the pro-

^{18&}lt;sub>Ibid.</sub>, p. 416.

¹⁹E. S. Pearson, "The Probability Integral of the Rangein Samples of Observations from a Normal Population," Biometrika, XXXII (1942), Parts 3 and 4, 301-7.

names $1_p = (w_p - \overline{w})/\sigma_w$, where w_p represents the range value corresponding to any one of the ordinates, having calculated the 1_p 's, it was only necessary to invert the rormula to obtain w_p from

wp = w + 1p ow.

restron used values of w as com uted by Tappett, but own

n = 2, 3, 6, 5, 6, 10, 20, 60, 100.

te bedugmon erew seulev Lenoitluba estaT

n = 30, 45, 75

by the same process of cubsture as that amployed by Tippett. From this framework the intermediate values of Twee obtained and finally the values of ip given in Pearson's Table A. 18 Since the values of w and Twee for samples drawn from a normal population, this table gives the percentage limits for the distribution of rance in samples from a normal population.

in a more recent article Pearson 9 stated that no simple express on exists for the probability law \$\darksquare\darksquar

¹⁸¹¹¹d., p. 416.

^{191.} S. Petraon, "The Probability Integral of the Adngeila Samples of Oceavetions from a Normal Population," Blometrika, TXXII (1942), Parts 3 and 4, 301-7.

bability integral, which gives the chance that the range in a sample of n observations is less than a given multiple of the population standard deviation. These values are more accurate than those previously obtained by Pearson and from them any confidence limits can be obtained. Hartley²⁰ showed the derivation of the complicated formula used and described the numerical evaluation of the probability integral which was accomplished by means of quadrature formulas and machine calculation.

IV. CONCLUSION

From this investigation and discussion it has been concluded that the sampling distribution of the range is asymmetrical but approaches most nearly to normal when 6 < n < 10. The method of moments offered the most satisfactory approach to the study of this distribution, with the moments calculated from the general equations with no assumptions as to linearity of regression or homoscedasticity of the correlation surface. Confidence limits were best obtained by computing the values of the probability integral according to values of the range and the size of the sample. All constants of the distribution of range were given in terms of the standard deviation of the population.

²⁰H. O. Hartley, "The Range in Random Samples," Biometrika, XXXII (1942), Parts 3 and 4, 341-42.

bacility integral, which gives the chance that the range in a sample of n observations is less than a given multiple of the population standard deviation. These values are more accurate than those previously obtained by Fearaon and from them any confidence limits can be obtained. Hartley 20 showed the derivation of the complicated formula used and described the numerical evaluation of the production formulas and machine was accomplished by means of quadrature formulas and machine calculation.

IV. COLULIUSION

From this investigation and discussion it as been concluded that the sampling distribution of the range is asymmetrical but approaches most perry to normal when b <n <10. The method of moments offered the most satisfactory approach to the study or this distribution, with the moments osloutated from the general equations with no assumptions as to linearity of regression or nomescediaticity of the correlation surface. Confidence limits were best obtained by computing the values of the probability integral according to values of the distribution of the probability integral according constants of the distribution of range were given in terms of the standard deviation of the population.

²⁰ n. O. Hartley, "The Hange in Random Samples," Blometrike, XXII (1942), Prts 3 and 4, 341-42.

PART II

APPLICATION OF RANGE TO QUALITY CONTROL

It has been previously stated that the range finds its principal use in its application to quality control. Specifically it is used in control chart analysis. In this respect the range can be employed to estimate the standard deviation of the population for control chart purposes or the control chart may be made for the ranges instead of for the standard deviations of the sample.

I. QUALITY CONTROL TECHNIQUE

Nature of quality control. The purpose of this paper is not to discuss quality control. Nevertheless, the subject of range cannot be adequately treated without some explanation of it. The idea of control involves action for the purpose of achieving a desired end, for example, detecting causes of trouble in processes, securing conformity with specifications, estimating the quality of a product, or the like. A manufacturer wishes to control a certain quality characteristic in his product. He cannot maintain an entirely uniform product. Yet if each factor which might cause variation in this characteristic continues during the process of manufacture to have the same probability of contributing a given effect, then the particular quality may

PART II

APPLICATION OF HAMES TO QUALITY CONTROL

It has been previously stated that the range finds its principal use in its application to quality convrol. Specifically it is used in control chart enalysis. In this respect the range can be employed to estimate the standard deviation of the population for control chart purposes or the control chart may be made for the ranges instead of for the standard deviations of the sample.

I. QUALITY CONTROL TECHNIQUE. I

Nature of quality control. Nevertheless, the subject is not to disques quality control. Nevertheless, the subject of range cannot be adequately treated without some explanation of it. The idea of control involves action for the purpose of achieving a desired and, for example, detecting causes of trouble in processes, securing conformity with specifications, estimating the quality of a product, or the line. A manufacturer wishes to control a certain quality characteristic in his product. He cannot maintein an entirely unitors product. He cannot maintein and entirely unitors product. Yet if each factor which sight the process of manufacture to have the same probability of the process of manufacture to have the same probability of contributing a given effect, then the particular quality may

be said to be controlled in a technical sense. 21

If the cost of testing the entire output is prohibitive, or if the test is destructive, it will be necessary to resort to sampling in order to determine whether uniformity of quality is being maintained. This involves statistical control. A standard level must be set and the limits within which the measurements of the quality characteristic may fluctuate without on the average departing from this level must be determined. The statistics for the successive samples must be recorded and comparisons made with the control limits.

The control chart. Now the control chart is a graphic representation of this type of data. On it are pictured the central value, or quality level desired, and the upper and lower control limits, that is, the boundaries within which the measures must remain if a state of control is to be maintained. When the findings have been plotted from sample to sample, the character of the output can be seen at a glance. Whether the quality of the product is unsatisfactory because the level of control does not

Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc. 1941), p. 348.

²²<u>Ibid</u>., p. 349.

be said to be controlled in a technical sense. 21

If the cost of testing the entire output is prohibitive, or if the test is destructive, it will be necessary to resort to sempling in order to determine whether uniformity of quality is being unintained. This involves statistical centrol. 22 A standard level must be set and the limits within which the measurements of the quality characteristic may fluctuate without on the average depicting from this level must be determined. The statistics for the successive samples must be recorded and domp risons made with the con-

The control chart. Now the control chart is a graphic representation of this type of data. On it are pictured the central value, or quality level desired, and the upper and lower control limits, that is, the boundaries within which the measures must remain if a state of control is to be maintained. When the findings have been plotted from comple to sample, the eneracter of the output can be seen at a glance. Whether the quality of the product is unsatisfactory because the level of control does not

Clrederick E. Groxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc. 1941), p. 348.

²² bld., p. 319.

remain constant, that is changes from one consignment to another, or because the level maintained, though constant, is too low, either because the mean of a variable characteristic is too small or because the standard deviation is too large --- in either case there is a type of control chart to suit the purpose. Changes in the level of control are most likely to be detected by an examination of the control chart for the means. But even when the level of control is maintained, individual units may fall below specification, if the variation within the sample is large. Satisfactory quality as well as a uniform product can be assured by the employment of a control chart for some measure of variability, such as the standard deviation. Other measures, especially the range, may be substituted for the standard deviation in such a chart.

Regardless of the form of chart used the standard deviation of the population is needed to set the control limits. If past experience has provided a standard of quality, for example, the population mean and standard deviation are known, the problem will simply be to discover whether fresh material continues to conform uniformly to this standard. If, on the other hand, no standard has been fixed, the standard deviation of the population must first

remain constant, that is onanges from one consignment to snother, or because the level maintained, though constant, is too low, either because the mean of a variable characteristic is too small or because the standard deviation is too large --- in either case there is a type of control chart to suit the purpose. Changes in the level of control are most likely to be detected by an examination of the control chart for the means. But even when the level of control trol is maintained, individual units may fall below specification, if the variation within the sample is large. Sattastory quality as well as a uniform product can be assured by the employment of a control chart for some desure of variability, such as the standard deviation.

Other measures, especially the range, may be substituted for the standard deviation in such a chart.

Regardless of the form of chart used the standard deviation of the population is needed to set the control limits. If past experience has provided a standard quality, for example, the population mean and standard daviation are known, the problem will simply be to discover whether fresh material continues to conform uniformly to this standard. If, on the other hand, no standard has been fixed, the standard deviation of the population must first

be determined.²³

Methods of determining $\underline{\sigma}$. E. S. Pearson²⁴ has given three methods of estimating this statistic:

(1) From the mean value of the variance:

$$G_c^2 = \frac{n}{N-k} (s_1^2 + s_2^2 + \cdots + s_k^2),$$

where N is the total number of observations, k the number of subgroups, n the number of units in each subgroup. σ_c will be the square root of the result.

(2) From the mean value of the standard deviation:

 $\sigma_{\text{C}} = \frac{1}{b_{\text{n}}} \cdot \frac{1}{k} (s_{1} + s_{2} + \cdots + s_{k}),$ where the factor, b_{n} , is the ratio of the mean value of the standard deviation of the samples to the standard deviation of the population for samples of size n. Pearson gives tables for b_{n} and $1/b_{\text{n}}$ for values of n from 2 to 30.

(3) From the mean value of the range:

$$\sigma_c = \frac{1}{dn} \cdot \frac{1}{k} (w_1 + w_2 + \cdots + w_k),$$

where d_n is the ratio of the mean range to the standard deviation of the population for samples of size n. Pearson gives

²³E. S. Pearson, The Application of Statistical Methods to Industrial Standardisation and Quality Control (London: British Standards Institution, 1935), p. 82.

^{24&}lt;sub>Ibid.</sub>, pp. 83-84.

be determined. 23

Methods of determining 6. E. S. Pearson²⁴ has given three methods of estimating this statistic:

(1) From the mean value of the variance:

$$\sigma_0^2 = \frac{n}{N-k} \quad (s_1^2 + s_2^2 + \dots + s_k^2),$$

where N is the total number of observations, k the number of subgroups, n the number of units in each subgroup.

of will be the square root of the result.

(2) From the mean value of the standard deviation:

Where the factor, b_n , is the ratio of the mean value of the standard deviation of the samples to the standard deviation of the population for samples of size a. Pearson gives tables for b_n and $1/b_n$ for values of a from a to a0.

 $oc = \frac{1}{dn} \cdot \frac{1}{1} (w_1 + w_2 + \dots + w_K),$

where do is the ratio of the mean range to the standard devia-

EbodieN leadisting to molication of Statistical Methods to Industrial Standards Institution, 1935, p. 82.

²⁴ Tuld., pp. 83-84.

tables for d_n and $1/d_n$ for values of n from 2 to 12. Tippett²⁵ has given full tables for this ratio for samples of between 2 and 1000 individuals drawn from a normal population. An abstract of these tables is given in Table II.

Part I. Since the mean range in these tables is in terms of the standard deviation of the original population, the latter can, in a particular case, be found by taking samples, determining the mean range, and dividing by the value given in the tables. The method is similar to one given by K. Pearson ²⁶ in which the sample is ranked and the difference between two certain individuals, preferably those near the quindeciles (those n/15 from each end), measured and divided by the value for a population having unit standard deviation.

From Table II it is seen how much the range depends on the size of the sample. Freeman²⁷ stated that for small k, say less than 10, or large n, say greater than 15, the mean range method of estimating of is unreliable, but that

²⁵L. H. C. Tippett, "On the Extreme Individuals and the range of Samples Taken from a Normal Population," Biometrika, XVII (1925), Parts 3 and 4,p. 386.

²⁶K. Pearson, (Editorial) "On the Probable Errors of Frequency Constants," Biometrika, XIII (1921), 113-119.

John Wiley & Sons, Inc., 1942), p. 131. (New York:

tables for dn and 1/dn for values of a from 2 to 12. Tippett²⁵ has given full tables for this ratio for sa [3ee of between 2 and 1000 individuals drawn from a normal population. An abstract of these tables is given in Table II.

This is the table of mean reages referred to in Part 1. Since the mean reage in these tables is in terms of the standard deviation of the original population, the latter can, in a particular case, be found by taking asuples, determining the mean range, and dividing by the value given in the tables. The method is similar to one given by h. Pearson 26 in which the sample is reaked and the difference between two certain individuals, preferably those near the quindeciles (those m/1; from each end), measured and divided by the value for a copulation maying unit standard deviation.

From Table II it is seen how much the range depends on the size of the sample. Treemon²7 stated that for small k, ear less than 10, or large n, say greater than 15, thu meen range method of estimating of is unreliable, but that

^{251.} H. G. Tippett, "On the Extreme Individuals and the range of Sangles Taken from a Mormal Population," Big-metriks, MVII (1925), Parts 3 and 4,p. 365.

Proquency Constants," Biometriks, Mil (1921), 113-119.

John Wiley & Sons, Inc., 1952), p. 131.

TABLE II

RATIO OF MEAN RANGE TO STANDARD DEVIATION

No. in Sample	Mean Range Standard Deviation
Dippost25 resumented coll	ecting the data into groups of 5
or the units 2	1.128
4	2.059
5	2.326
10	3.078
50	4.498
100	5.015
500	6.073
state of control. Beverth	

TABLE II

RATIO OF MEAN RANGE TO STANDARD DEVIATION

Mean Range Standard Deviation	No. in Sample		
1.128			
2.059			
2.326			
3.078			
4.498			
5.015			
6.073	500		

for n less than 15 a good estimate of σ can be formed from the mean range. Pearson²⁸ said that no estimate of σ can be considered satisfactory if the total number of observations is less than 30, but that if N (= nk) is greater than 50 and the observations have been broken up into equal sub-groups each containing not more than 10 units, the estimate will be adequate for control chart purposes. Tippett²⁹ recommended collecting the data into groups of 5 or 10 units.

II. APPLICATION OF THE TECHNIQUE

The <u>sample</u>. A problem has been considered as best illustrating the approximations of obtained by the three methods suggested. Shewhart³⁰ has given a set of 204 observations of the measurements in megohms of the insulation resistances of as many different pieces of a new kind of material produced under presumably the same essential conditions. The particular characteristic measured was not in a state of control. Nevertheless, the sample will serve for

E. S. Pearson, The Application of Statistical Methods, p. 84.

²⁹L. H. C. Tippett, The Methods of Statistics (second edition, revised; London: Williams and Norgate, Ltd., 1937) p. 32.

³⁰Walter A. Shewhart, Statistical Method from the View-point of Quality Control (Washington: U. S. Department of Agriculture, 1939), p. 90.

for n less than 15 a good estimate of o can be formed from the mean range. Pearson 28 said that no estimate of o can be considered satisfactory if the total number of observations is less than 30, but that if N (= nk) is greater than 50 and the observations have been broken up into equal sub-groups each containing not more than 10 units, the estimate will be dequate for control chart purposes.

Tippett 29 recommended collecting the deta into groups of 5 or 10 units.

II. AFFICATION OF THE TECHNIQUE

The sample. A problem has been considered as best illustrating the approximations of a obtained by the digree methods suggested. Shewhart of has given a set of 204 observations of the measurements in megohms of the insulation resistances of as many different pieces of a new kind of meterial produced under presumably the same essential conditions. The perticular characteristic researed was not in a state of control. Nevertheless, the sample will serve for

Listhode, p. 84.

ealtion, revised; London: Williams and Worgete, 104., 1937)

Journter .. Chevantt, Statistical Method from the Viewpoint of Quality Control (Washington: U. S. Department of
Agriculture, 1939), p. 90.

illustrative purposes. Table III gives these measurements in the order in which they were taken.

Statistics of the sample. The sample has been divided into 51 subgroups of 4 units each, the order in which the measurements were taken being preserved, since that is the order which should furnish the clue to assignable causes of variability. The means (\overline{x}) , variances (s^2) , standard deviations (s), and ranges (w) of these subgroups are given in Table IV.

Computation of _____. Substitution of the values from Table IV in the formulas above gives the following results:

For
$$N = 204$$
, $n = 4$, $k = 51$

From
$$s_1^2$$
 $C_c = \sqrt{\frac{4}{153} \cdot 4834406} = 355.5$

From
$$s_i$$
 $c = 1.253 \cdot \frac{1}{51} \cdot 13358 = 328.2$

From wi
$$G_c = 0.4857. \frac{1}{51}$$
 . 33600 = 320.0

Control limits. If the control chart is made for the means, the estimate from the mean range is sufficiently accurate and a distinct time-saver. If, on the other hand, the control chart is made for the standard deviations, the estimate from the mean standard deviation or from the mean variance is more in order, so as to eliminate any unnecessary

illustrative purposes. Table III gives these measurements in the order in which they were taken.

Statistics of the sample. The sample has been divided into 51 sungroups of 4 units each, the order in which the measurements were taken being preserved, since that is the order which should furnish the olde to assignable causes of variability. The means (\vec{x}) , variances (s^2) standard devictions (s), and ranges (w) of these subgroups are given in Table IV.

Computation of 6 Substitution of the values from Table IV in the formulas above gives the following results:

For M = 204, n = 4, k = 51

Control limits. If the control chart is made for the mean, the estimate from the mean range is sufficiently accurate and a distinct time-saver. If, on the other hand, the control chart is made for the standard deviations, the estimate from the mean standard asylation or from the mean variance is more in order, so as to eliminate any unnecessary

TABLE III

OBSERVATIONS OF 204 MEASUREMENTS OF INSULATION RESISTANCE

				-		
5045 4350 4350 3975	4790 4845 4700 4600	4090 5000 4335 5000	5000 4575 4700 4430	4840 4310 4185 4570	5000 4700 4500 4840	4625 4425 4135 4190
4290 4430 4485 4285	4100 4410 4180 4790	4640 4335 5000 4615	4850 4850 4570 4570	4700 4440 4850 4125	5075 5000 4770 4570	4080 3690 5050 4625
3980 3925 3645 3760	4790 4340 4895 5750	4215 4275 4275 5000	4855 4160 4325 4125	4450 4450 4850 4450	4925 4775 5075 4925	5150 5250 5000 5000
3300 3685 3463 5200	4740 5000 4895 4255	4615 4735 4215 4700	4100 4340 4575 3875	3635 3635 3635 3900	5075 4925 5250 4915	
5100 4635 5100 5450	4170 3850 4445 4650	4700 4700 4700 4095	4050 4050 4685 4685	4340 4340 3665 3775	5600 5075 4450 4215	
4635 4720 4810 4565	4170 4255 4170 4375	4095 3940 3700 3650	4430 4300 4690 4560	5000 4850 4775 4500	4325 4665 4615 4615	
4410 4065 4565 5190	4175 4550 4450 2855	4445 4000 4845 5000	3075 2965 4080 4080	4770 4500 4770 5150	4500 4765 4500 4500	
4725 4640 4640 4895	2920 4375 4375 4355	4560 4700 4310 4310	4425 4300 4430 4840	4850 4700 5000 5000	4850 4930 4700 4890	570

OBSERVATIONS OF 204 MEASUREMENTS OF THRUTATION RESISTANCE

5244	1,700	4310				6350
56741						
						3975
			4850	0404	4100	4290
				4335	14410	
5050		14850	1,570		0814	4465
4625					4790	1285
5150		4450	4855			
5250	4775					
	2114	4450		4275	4340	
		7820	4325		4895	3645
					5000	
					4255	
						4635
	0577	3005			4444	2100
					M170	4635
					4255	
	4615				4170	
	ALLE					
			0954		4375	4565
	4705					
		4770				
		5150	0804			
			0004			0615
		4,850	4425			
	16930	4700			4375	
	1700	5000			4375	
		5000			1,355	1895
					coin	

TABLE IV

STATISTICAL MEASURES FOR DISTRIBUTION OF INSULATION RESISTANCE MEASUREMENTS DIVIDED INTO 51 SUBGROUPS OF 4 UNITS EACH

Sample No.	₹	s ²	S	w	
1	4430	149512	387	1070	
2	4372	7606	87	200	
3	3828	17656	133	335	
4	3912	571654	756	1900	
5	5071	83855	290	815	
6	4682	8432	92	245	
7	4558	166106	408	1125	
8	4725	10838	104	255	
9	4734	8642	93	245	
10	4370	71750	268	690	
11	4944	260142	510	1410	
12	4722	81406	285	745	
13	4279	90280	300	800	
14	4242	7056	84	205	
15	4008	461606	679	1695	
16	4006	393380	627	1455	
17	4606	162542	403	910	
18	4648	55756	236	665	
19	4441	104667	324	785	
20	4566	43030	207	520	
21	4549	68630	262	605	
22	3846	32642	181	445	
23	4572	150256	388	1000	
24	4470	28050	167	390	
25	4676	44067	210	570	

MARION OF STREET BOY THEFT OF THE OFFICE OF INSULATION OF THE OFFICE OFFICE OF THE OFFICE OF THE OFFICE OF THE OFFICE OF THE OFFICE OFF

1				alt at
2				
2				
2	Tarar.	CRE IN		
1982 1986 1985				
\$ 71050 \$ 4900 \$ 15				
5				
1		e boles as		
7				
7				
1				
10				
10 4970 71750 240 700 1410 1410 1410 1410 1410 1410 141				
12				
12				
33				
33	915			
14				
15				W. Aller
16				
17				
17				
20 25750 255 005 20 2441 104667 324 765 20 257 258 20 257 258 22 257 258 23 257 258 24 257 258 257 258 258 258				
20 22 20 20 20 20 20 20 20 20 20 20 20 2				
26 20 207, 528 21 1019 202 005 22 202 2020 35062 131 445 23 35 35 20 205 388 100 24 45 40 205 107 200 25 35 36 40 600 107 200 25 36 40 600 700 700				
20 202 005 20 202 101 445 23 24 4572 1,0256 388 1000 24 4470 25050 127 230 25 4470 25050 127 230 25 4470 25050 127 230				
			The state of the	

Sample No.	<u>x</u>	s ²	S	W
26	4710	19600	140	280
27	4366	85330	292	730
28	4222	68456	262	700
29	4368	100806	317	635
30	4495	21125	145	390
31	3550	282412	531	1115
32	4499	41530	204	540
33	4476	63392	252	655
34	4529	75855	275	725
35	4550	30000	173	400
36	3701	13167	115	265
37	4030	97612	312	675
38	4781	32930	181	500
39	4798	53569	231	650
40	4888	15469	124	300
41	4760	33800	184	500
42	4854	39467	199	505
43	4925	11250	106	300
44	5041	18542	136	335
45	4835	293862	542	1385
46	4555	18050	134	340
47	4566	13167	115	265
48	4842	7569	87	230
49	4344	38230	196	490
50	4361	268405	518	1360
51	5100	11250	106	250
Totals		4834406	13358	33600

			-		
	W		S _B	X	Sample No.
			1125 100806 85330 100806	4710 4365 4222 4365 4365	
	111 50 72 72 74 74		282412 41530 75855 30000		32 32 32 32 32 32 32 32 32 32 32 32 32 3
900					
		136 136 136 136	33802 11250 39467 39467 33467		
0000	34 25 23 49 136 136		18050 7569 7569 38230 208405 11250	4555 4566 4842 4344 4361 4361	
			4834406		

computations. But here is an important point. Because of this relation between the range and the standard deviation a control chart made for the range exhibits the same variations as the control chart for the standard deviations.

Plotting the data of Table IV for standard deviation and for range, as in Figures 7 and 8, illustrated this fact.

E. S. Pearson³¹ has fixed the control limits for the standard deviation as follows:

Outer control limits at $B_{0.001}$. σ and $B_{0.999}$. σ Inner control limits at $B_{0.025}$. σ and $B_{0.975}$. σ Mean value of $s = b_n \sigma$.

The quantities B and b for samples of from 2 to 30 units are given in his text and are for n = 4: $B_{0.001} = 0.078$, $B_{0.025} = 0.232$, $B_{0.975} = 1.529$, $B_{0.001} = 2.017$. These values give an outer pair of limits within which (were the variation statistically uniform) 99.8 per cent of the values of s_i should fall and an inner pair within which 95 per cent should fall.

For the range the 95 per cent and 99.8 per cent limits are set up in a similar manner.

Methods, p. 88. The Application of Statistical

^{32&}lt;sub>Ibid.</sub>, p.86.

computations. But here is an important coint. Receive of this relation between the range and the standard deviction a control obset made for the range exhibits the seme veriations as the control chart for the standard deviations.

Flotting the data of Table IV for standard deviation and for range, as in Figures 7 and 8, illustrated this fact.

S. Fearson 1 has fixed the control limits for the standard deviation as follows:

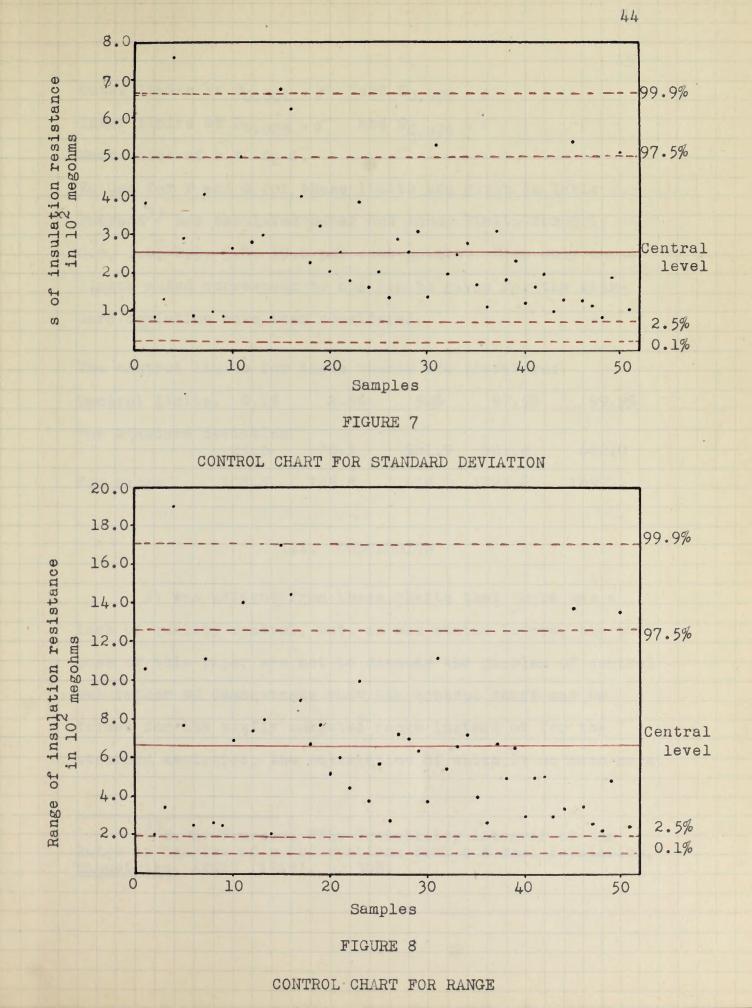
Outer control limits at Bo.025.0 and Bo.995.0

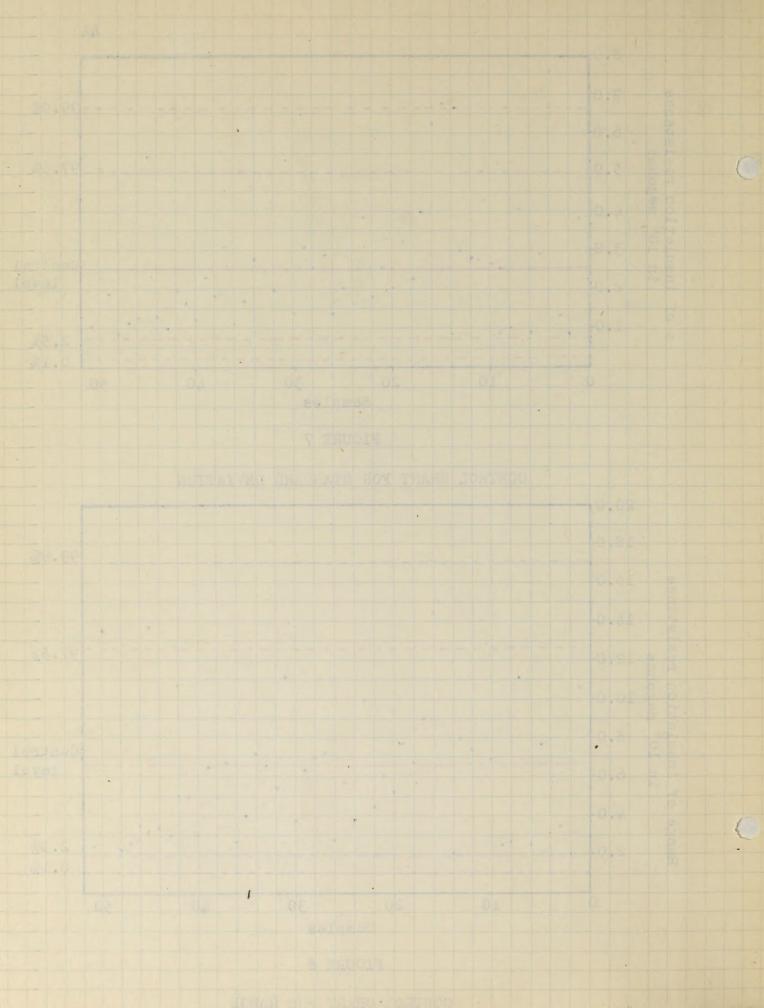
The quantities 2 and b rot samples of from 2 to 30 units are given in his text and are for n=4: $B_{0.001}=0.078$, $B_{0.025}=0.232$, $B_{0.975}=1.529$, $B_{0.001}=2.017$. These values give an outer pair of limits within which (were the variation statistically uniform) 99.3 per cent of the values of si should fall and so inner pair within which 95 per cent should fall.

For the range the 95 per cent and 99.8 per cent limits are set up in a similar manner.

Mr. S. Pestson, The Application of Statisticsl

^{32 101}d., p.85.





Outer limits at $D_{0.001} \cdot \sigma$ and $D_{0.999} \cdot \sigma$ Inner limits at $D_{0.025} \cdot \sigma$ and $D_{0.975} \cdot \sigma$ Mean value of $r = d_n \sigma$.

Values for D and d for these limits are given in Table V.

Pearson³³ has tabulated outer and inner limits for 0.1,

0.5, 1.0, 2.5, 5.0, 10.0 per cent points. Here only those

limits which correspond to the limits given for the standard deviation have been considered.

The control limits for these charts are therefore:

Control limits. 0.1% 2.5% 50% 97.5% 99.9%

For standard deviation 25.6 76.1 261.9 501.8 662.0

For range 108.8 188.8 658.9 1273.6 1699.2

III. CONCLUSION

It was evident from these charts that there was a lack of quality control, but, as was stated before, the purpose of this paper was not to discuss the problem of control but rather to demonstrate that the control chart may be formed for the easily computed range instead of for the standard deviation, the calculation of which is so much more

³³E. S. Pearson, "The Probability Integral of the Range in Samples of n Observations from a Normal Population," Biometrika, XXXII (1942), p. 308.

Outer limits at Do. 001 . 6 and Do. 999 . 6

loser thatte at Do. 025 . o and Do. 975 . o

Values for D and a for these limits are given in Table V. Pearson in table to D.1, - 0.5, 1.0, 2.5, 5.0, 10:0 per cent points. Here only those limits which correspond to the limits given for the standard deviation have been considered.

The control limits for where durits are therefore:

Gontrol limits. 0.11 2.5% 50% 97.5% 99.9%

For avandard deviation 25.6 70.1 261.9 501.8 662.0

For range 103.6 186.8 658.9 1273.6 1699.2

III. CONCLUSION

Lt was evident from these charts that there was used and of quality control, out, as was stated before, the purpose of this paper was not to discuss the problem of control out rather to demonstrate that the control onert may be formed for the easily computed renge instead of for the standard deviction, the calculation of which is so much more

Hange in Samples of n Chestyettons from a normal Topulation," Biometrika, XXXII (1942), p. 306.

TABLE V

PROBABILITY LIMITS FOR DISTRIBUTION OF RANGE

Size Sampl		ver percen	_	s Upper percent Do.975	tage points D _{0.999}
2	0.8862	0.00	0.04	3.17	4.65
3	0.5908	0.06	0.30	3.68	5.06
4	0.4857	0.20	0.59	3.98	5.31
5	0.4299	0.37	0.85	4.20	5.48
6	0.3946	0.54	1.06	4.36	5.62
7	0.3698	0.69	1.25	4.49	5.73
8	0.3512	0.83	1.41	4.61	5.82
9	0.3367	0.96	1.55	4.70	5.90
10	0.3249	1.08	1.67	4.79	5.97
11	0.3152	1.20	1.78	4.86	6.04
12	0.3069	1.30	1.88	4.92	6.09
1	a range has b	went bruve	n theorati	cally ar two w	

TABLE V.

PROBABILITY LIMITS FOR DISTRIBUTION OF HAMCE

Do.999	Upper perces	ntage points Do.025		dn	
4.65	3.47	40.0	0.00	0.8862	
5.06	3.68	0.30	0.06	0.5908	
5.31	3.98	0.59	0.20	0.4857	
5.48	05.4	28.0		0.4299	
50.62	4.36	1.00	0.54	0.3946	
5.73	4.49	1.25	0.69	0.3698	
5.82	4,61	1.41	0.83	0.3512	
5.90	4.70	1.55	0.96	0.3367	
5.97	4.79	1.67	1.08	0.3249	
40.0	4.86	1.78	1.20	0.3152	II
60.0	58.3	1.88	1.30	0.3069	

laborious, and that such a chart is not less meaningful. A study of the charts revealed the similarity in pattern between the two, which indicates that the simpler chart may be usefully employed for control purposes. This conclusion has been confirmed by theoretical investigation.

However, Pearson called attention to the following facts regarding the range:

- (1) Apart from these special applications when dealing with a number of small groups, the use of the range is to be deprecated since it provides a far less accurate measure of variation than that given by the standard deviation.
- (2) Its use in control chart analysis can only be recommended where each sample contains not more than 10 units, since as n increases the range becomes a less and less reliable measure of variation, depending only on the extreme values and taking no account of the form of variation between these.
- (3) The condition that the variation due to chance causes should be of the normal form is somewhat more stringent than in the case of charts for standard deviation. 34

Nevertheless, in spite of these apparent handicaps, the range has been proven theoretically to be a very useful measure. To some extent it has already been put to practical use and promises in future, once its nature is more thoroughly understood, to have even wider application.

³⁴ E. S. Pearson, The Application of Statistical Methods, p.89.

laborious, and that such a chart is not less seningful. A study of the charts revealed the similarity in pattern between the two, which indicates that the similar chart may be usefully employed for control purposes. This conclusion has been confirmed by theoretical investigation.

However, Pearson called strention to the following

- (1) Apart from these elected applications when desling with a number of small groups, the use of the range le to be degreeated since it provides a far less anourate measure of variation than that given by the stendard deviation.
 - (a) its use in control enert analysis can only be recommended where each sapple contains not core than 10 units, since as n increases the range becomes a less and less relisole measure of variation, depending only on the extress values and taking no account of the form of variation between these.
 - dauses should be of the northeter tor assess and the case of cherts for standard deviastructure of the case of cherts for standard devia-

devertheless, in spite of these apparent handicaps, the range has been proven theoretically to be a very userful messure. To some extent it has already been but to predicted use and promises in future, once its nature is more thoroughly understood, to have even wider application.

Methods, p.89.

BIBLIOGRAPHY

- Croxton, Frederick E. and Dudley J. Cowden, Applied General Statistics. New York: Prentice-Hall, Inc. 1941. 944 pp.
- Fisher, R. A. and L. H. C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest and Smallest Members of a Sample," Proceedings of the Cambridge Philosophical Society, XXIV (1928), Part 2, 180-90.
- Freeman, H. A., Industrial Statistics. New York: John Wiley & Sons, Inc., 1942. 178 pp.
- Gumbel, E. J., "Ranges and Midranges," Annals of Mathematical Statistics, XV (1944), 414-22.
- Hartley, H. O., "Numerical Evaluation of the Probability Integral," Biometrika, XXXII (1942) Parts 3 and 4, 309-10.
- "The Range in Random Samples," Biometrika, XXXII (1942), Parts 3 and 4, 334-48.
- Pearson, E. S., "A Further Note on the Distribution of Range in Samples Taken from a Normal Population," Biometrika, XVIII (1926), 173-94.
- Standards Institution, 1935. 161 pp.
- , "The Percentage Limits for the Distribution of Range in Samples from a Normal Population," Biometrika, XXIV (1932), 404-17.
- of n Observations from a Normal Population," Biometrika, XXXII (1942) Parts 3 and 4, 301-8.
- Pearson, Karl, "Note on Francis Galton's Problem," Biometrika I (1902), 390-99.
- Constants." Biometrika, XIII (1921), 113-32.

BUBILL OCH LEAY

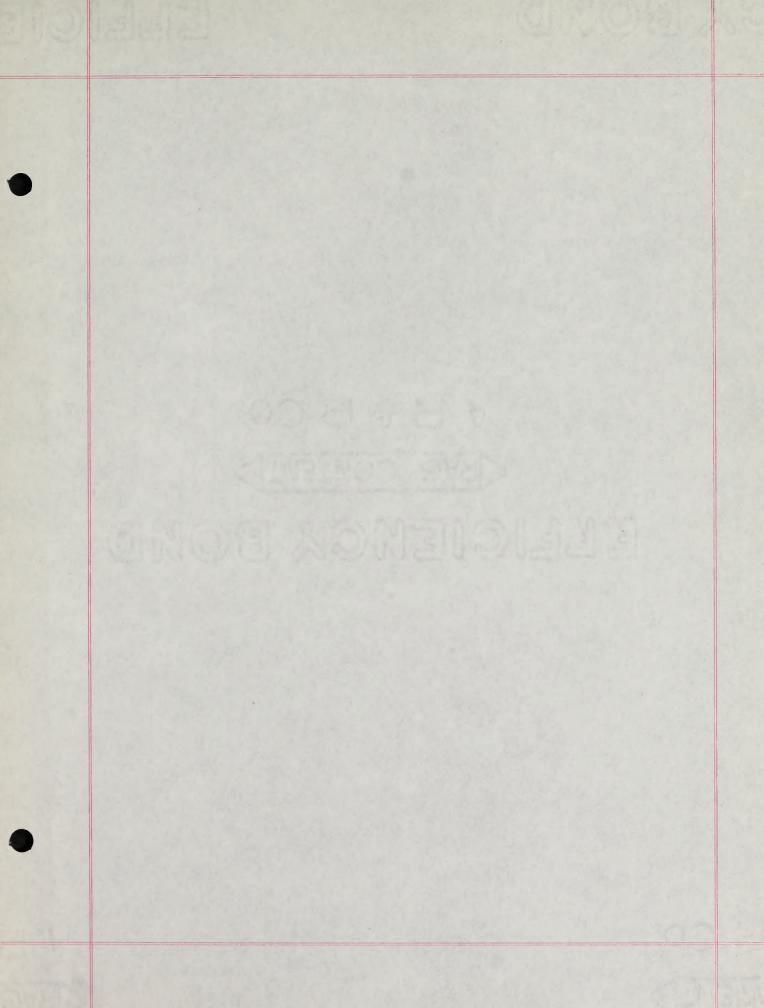
- Croxton, Frederick E. and Dugley J. Cowden, Applied General Statistics, New York: Prentice-Hall, Inc. 1941. 944 pp.
 - Figure, R. A. and I. H. C. Tippett, "Limiting Forms of the frequency Distribution of the Largest and Smallest Members of a sample," Proceedings of the Combridge Philosophical Society, NAIV (1928), Part 2, 180-90.
- Froeman, H. A., Inquatrial Statistics. New York: John Wiley & Sons, anc., 1942. 178 pp.
- Gumbel, E. J., "Ranges and Midranges," innals of Mathematical Statistics, AV (1944), 414-22.
 - Hartley, H. C., "Numerical system of the frobability integral," Biometrika, XXXII (1942) Parts 3 and 4, 309-10.
 - "The Range in Random Samples," Blometriks, XXXII (1942), Perts 3 and 6, 336-48.
- Pearson, E. S., "A partner Note on the Distribution of Range in Scaples Taken from a Normal Population," Biomotrike, AVIII (1926), 173-94.
- Standards institution, 1935, 101 pp. Standards artish Standards institution, 1935, 101 pp.
- "The Fercentage Limits for the Distribution of Mange in Samples from a Mormal Population," Biometrike,
- of a Observations from a Normal Population," Biometrike,
- Pearson, Mari, "Note on Francis Galton's Froblem," Biometrika
 - "Onabanta." Biometrika, Mar (1921), 113-32.

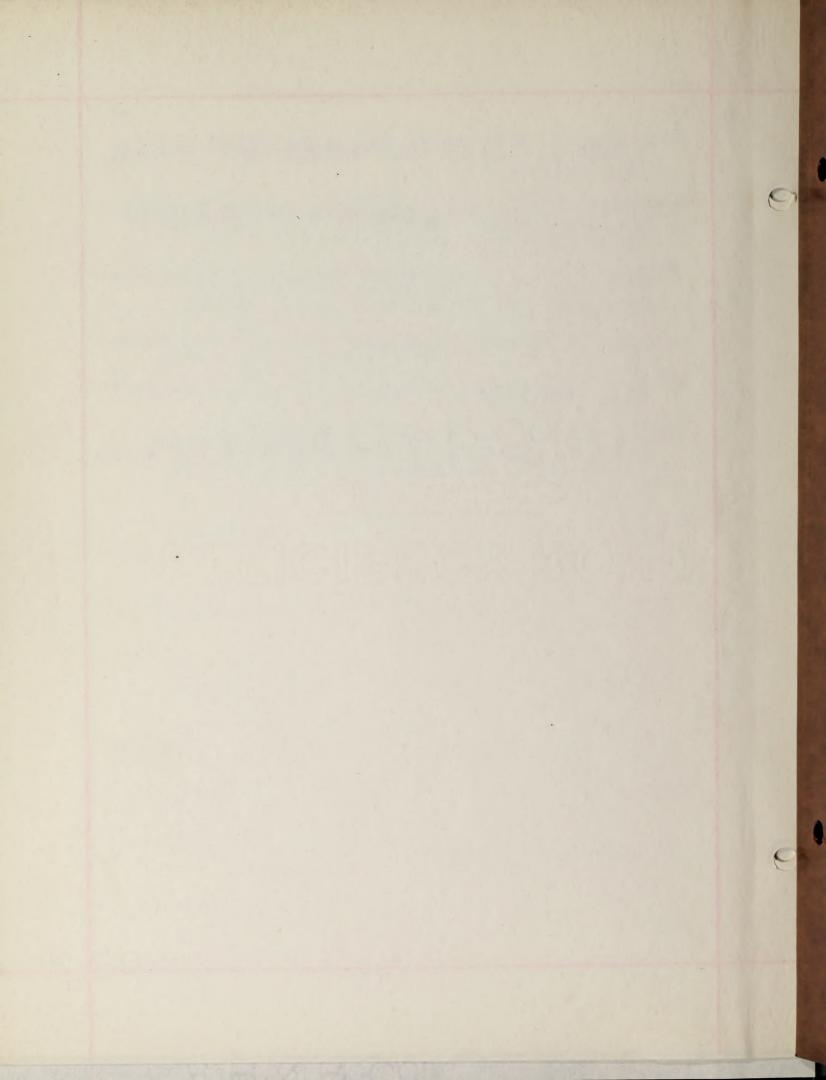
- Richardson, C. H., An Introduction to Statistical Analysis, Enlarged edition; New York: Harcourt, Brace and Company 1941. 312 pp.
- Shewhart, Walter A., Statistical Method from the Viewpoint of Quality Control. Washington: U.S. Department of Agriculture, 1939. 155 pp.
- Tippett, L. H. C., "On the Extreme Individuals and the Range of Samples Taken from a Normal Population,"

 Biometrika, XVII (1925), Parts 3 and 4, 364-87.
- London: Williams and Norgate, Ltd., 1937. 280 pp.
- Wilks, S. S., <u>Mathematical</u> <u>Statistics</u>. Princeton University Press, 1944. 284 pp.
- Yule, G. Undy and M. G. Kendall, An Introduction to the Theory of Statistics. Eleventh edition, revised; London: Charles Griffin and Company, Ltd., 1937. 570 pp.

- Richardson, C. H., An Introduction to Statistical Inalysis, Malerged edition; New York: Hardourt, Brace and Company 1941. 312 pp.
 - Shewhart, Walter A., Statistical Method from the Viewpoint of Juelity Control. Washington: U. S. Department of Agriculture, 1939. 155 pp.
- Tippett, L. H. C., "On the intreme Individuals and the Range of Samples Taken from a Normal Population."

 Blometrika, XVII (1925), Parts 3 and 4, 364-87.
- The Methods of Stetletlos. Second edition, revised; London: Williams and Norgate, Ltd., 1937. 200 pp.
- Wilks, S. S., Mathematical Statistics. Princeton University iress, 1944. 284 pp.
- Yule, G. Undy and M. G. Kendall, an introduction to the Theory of Statistics. Eleventh cultion, revised; London: Charles Griffin and Company, itd., 1937. 570 pp.







CCOPRESS BINDER

BES 250-P7

MADE BY

CCO PRODUCTS, INC.

ING ISLAND CITY, N.Y., U. S. A.

